Scriptum zur Vorlesung
Quadratische Formen

Prof. W. Hoffmann
SS 2010

Quadratische Formen, der Gegenstand dieser Vorlesung, sind Polynome mit bestimmten Eigenschaften. Darum werden wir zunächst den Begriff des Polynoms genauer betrachten.

1 Polynome

Verknüpft man Zahlen durch Rechenoperationen, so entstehen Rechenausdrücke, auch Terme genannt. Wiederholt sich derselbe Rechenweg mit verschiedenen Ausgangswerten, so bildet man Rechenausdrücke mit Variablen, für die dann die Ausganswerte eingesetzt werden können, zum Beispiel

πr^2	Flächeninhalt eines Kreises mit dem Radius r
$\gamma\dfrac{m_1 m_2}{r^2}$	Gravitationskraft zwischen zwei Massen der Größe m_1 und m_2 im Abstand r
$n!$	Anzahl der Permutationen von n Dingen
$\sqrt{a^2+b^2}$	Länge der Diagonale in einem Rechteck mit den Seitenlängen a und b

Hier sind r, m_1, m_2, n, a und b Variablen, auch Unbestimmte genannt, während π und γ feste Zahlen sind, auch Konstanten genannt. (Wir übergehen der Einfachheit halber die Tatsache, dass zur Angabe von physikalischen Größen neben Zahlen auch noch Maßeinheiten gehören.)

Polynome sind grob gesprochen solche Terme, in denen nur die Operationen Addition und Multiplikation vorkommen, wie z. B.

$$\pi rr, \qquad (3xx+2xy)(5x+yy+4y), \qquad (5xx+yxy+4xy)(2y+3x).$$

Terme, in denen Potenzen mit natürlichen Exponenten sowie Differenzen vorkommen, sind auch zugelassen, da wir z. B. r^2 als rr und $a-b$ als $a+$

$(-1)b$ schreiben können. Wenn keine Verwechslungen zu befürchten sind, lässt man das Multiplikationszeichen meist weg. Mehrfache Produkte oder Summen ohne Klammersetzung werden von links nach rechts berechnet. Von den eingangs genannten Termen ist übrigens nur der erste ein Polynom.

Die letzten beiden Ausdrücke ergeben immer den gleichen Wert, egal welche Zahlen man für x und y einsetzt, weil man sie unter Benutzung der Rechengesetze ineinander umformen kann:

$$\begin{aligned}(3xx+2xy)(5x+yy+4y) &\overset{\mathrm{K}^{\cdot}}{=} (3xx+2yx)(5x+yy+4y)\\ &\overset{\mathrm{D}}{=} (3x+2y)x(5x+yy+4y) \overset{\mathrm{D}}{=} (3x+2y)(x5x+xyy+x4y)\\ &\overset{\mathrm{K}^{\cdot}}{=} (5xx+yxy+4xy)(3x+2y) \overset{\mathrm{K}^{+}}{=} (5xx+yxy+4xy)(2y+3x).\end{aligned}$$

Hier bezeichnen wir mit K^{+} das Kommutativgesetz der Addition, mit $\mathrm{K}^{\cdot}$ das der Multiplikation und mit D das Distributivgesetz. Hätten wir noch sorgfältiger gearbeitet und bei allen Summen und Produkten die Reihenfolge der Berechnung durch Klammern angegeben, so hätten wir noch die Assoziativgesetze A^{+} und $\mathrm{A}^{\cdot}$ benötigt.

Definition 1. *Zwei Terme, in denen Zahlen und Variablen durch die Zeichen + und · verknüpft sind, nennen wir äquivalent, wenn man den Einen unter Benutzung der Rechengesetze für die Addition und die Multiplikation sowie durch die Ausführung von Rechenoperationen in den Anderen umformen kann. Die Äquivalenzklassen von solchen Termen nennen wir Polynome.*

Mit dem Ausführen von Rechenoperationen meinen wir auch, dass man das Produkt von 0 mit einem Term durch 0 ersetzen kann und dass man das Produkt von 1 mit einem Term durch diesen Term ersetzen kann. Üblicherweise führt man für die Äquivalenzklassen keine neue Bezeichnung ein, sondern benutzt das Gleichheitszeichen, um anzugeben, dass zwei Terme äquivalent sind.

Man kann jeden Term in eine äquivalente Standardform bringen. Solange in dem Ausdruck noch ein Produkt vorkommt, bei dem einer der Faktoren eine Summe ist, wandeln wir dieses mit Hilfe des Distributivgesetzes um:

$$\begin{aligned}(3xx+2xy)(5x+yy+4y) &= 3xx(5x+yy+4y)+2xy(5x+yy+4y)\\ &= 3xx5x+3xxyy+3xx4y+2xy5x+2xyyy+2xy4y.\end{aligned}$$

Auf diese Weise erhalten wir eine Summe von Produkten. In jedem Summanden fassen wir unter Benutzung des Kommutativgesetzes die Zahlenfaktoren zu einer Konstanten und gleiche Variablen zu Potenzen zusammen, wobei wir

uns an eine feste Reihenfolge der Variablen halten. Ein Produkt von Potenzen der Variablen nennt man Monom und die Konstante, die man taditionell davor schreibt, den Koeffizienten des Monoms. Schließlich fassen wir noch die Terme mit den gleichen Monomen mit Hilfe des Distributivgesetzes zusammen und berechnen die entstehenden Summen von Zahlen. In unserem Beispiel erhalten wir

$$15x^3 + 3x^2y^2 + 22x^2y + 2xy^3 + 8xy^2.$$

Kommt ein Monom nicht vor, wie hier z. B. xy, so können wir es mit dem Koeffizienten 0 hinzuaddieren. Steht vor einem Monom kein Zahlenfaktor, so können wir den Koeffizienten 1 davorschreiben. Kommt eine Variable in einem Monom nicht vor, so können wir sie mit dem Exponenten 0 dazuschreiben. Einen konstanten Summanden schreiben wir als Koeffizienten des Monoms, in dem jede Variable mit dem Exponenten Null auftritt.

Wenn wir die Klammern in anderer Reihenfolge auflösen, können sich dann die Koeffizienten der Monome ändern? Wie erkennt man, ob zwei Terme äquivalent sind? Hier sind die Antworten auf beide Fragen:

Satz 1. *Die Koeffizienten der Monome eines Termes wie in Definition 1 sind eindeutig bestimmt. Zwei Terme sind genau dann äquivalent, wenn ein beliebiges Monom in beiden den selben Koeffizienten hat.*

Der Beweis dieses Satzes erfordert weitere Begriffe. Wir werden später die Beweisidee andeuten.

2 Ringe

Bisher haben wir allgemein von Zahlen gesprochen. Für manche Anwendungen sind nur Zahlen aus einem bestimmten Zahlbereich sinnvoll. Bei der Berechnung eines Ausdrucks mit Zahlen aus einem gegebenen Zahlbereich wird auch das Ergebnis in diesem Zahlbereich liegen, wenn er unter den beteiligten Operationen abgeschlossen ist. In unserem Fall sind das die Addition und die Multiplikation. Eigentlich kommt es gar nicht darauf an, dass die Objekte, die wir addieren und multiplizieren, Zahlen sind.

Definition 2. *Ein Ring ist eine Menge R, auf der zwei Operationen $+$ und $\cdot$ mit folgenden Eigenschaften gegeben sind.*

(i) Für beliebige Elemente a, b und c von R gilt

$$a + b = b + a, \qquad (a + b) + c = a + (b + c),$$
$$a \cdot b = b \cdot a, \qquad (a \cdot b) \cdot c = a \cdot (b \cdot c),$$
$$(a + b) \cdot c = a \cdot c + b \cdot c.$$

(ii) Es gibt Elemente 0 und 1 von R, so dass für alle Elemente a von R gilt

$$0 + a = a, \qquad 1 \cdot a = a.$$

(iii) Für jedes Element a von R gibt es genau ein Element b von R, genannt das entgegengesetzte Element von a, so dass $a + b = 0$.

Häufig wird in der Definition eines Ringes die Kommutativität der Multiplikation und die Existenz des Einselements nicht gefordert. In dieser Vorlesung werden wir den Begriff des Ringes aber in dem eben definierten Sinn benutzen.

Beispiele für Ringe sind der Bereich **Z** der ganzen Zahlen, der Bereich **Q** der rationalen Zahlen und der Bereich **R** der reellen Zahlen. Hingegen sind die Bereiche der natürlichen Zahlen und der Bruchzahlen keine Ringe. Strenggenommen müsste man die Operationen Addition und Multiplikation sowie das Nullelement und und das Einselement für jeden Ring anders bezeichnen. Meist ist aber die Bedeutung der Symbole $+$, $\cdot$, 0 und 1 aus dem Zusammenhang klar.

Das einfachste Beispiel eines Ringes ist die Menge $\{0\}$. Hier ist 0 sowohl das Null- als auch das Einselement.

Als weiteres Beispiel eines Ringes betrachten wir die Menge $\{\mathrm{g}, \mathrm{u}\}$ mit den Operationen

$+$	g	u
g	g	u
u	u	g

$\cdot$	g	u
g	g	g
u	g	u

In diesem Ring ist g das Nullelement und u das Einselement.

Man kann sich leicht überzeugen, dass es in einem Ring nur ein Nullelelent geben kann. Ist nämlich $0'$ ebenfalls ein Nullelement, so gilt nach Eigenschaft (ii) und dem Kommutativgesetz der Addition $0' = 0 + 0' = 0' + 0 = 0$. Genauso zeigt man, dass es nur ein Einselement geben kann. Außerdem gilt für alle Elemente a von R, dass

$$0 \cdot a = 0.$$

Nach dem Distributivgesetz und Eigenschaft (ii) ist nämlich

$$0 \cdot a + 0 \cdot a = (0 + 0) \cdot a = 0 \cdot a.$$

Mit dem Assoziativgesetz folgt für alle Elemente b

$$(b + 0 \cdot a) + 0 \cdot a = b + (0 \cdot a + 0 \cdot a) = b + 0 \cdot a.$$

Nun genügt es, für b das entgegengesetzte Element von $0 \cdot a$ zu wählen.

Alles im vorigen Abschnitt über Zahlen gesagte lässt sich auf die Elemente eines Ringes R anwenden. Wir können also Terme betrachten, in denen Elemente von R und Variablen durch die Zeichen $+$ und $\cdot$ verknüpft werden. Man muss hier Variablen gut von den Buchstaben unterscheiden, die Ringelemente bezeichnen. Die Äquivalenz von Termen ist wie in Definition 1 erklärt, wobei jetzt die Ringoperationen zur Anwendung kommen. Äquivalente Terme ergeben den selben Wert, wenn wir für die Variablen Ringlelemente einsetzen und die mit $+$ und $\cdot$ notierten Operationen ausführen. Man spricht dann von Polynomen mit Koeffizienten in R.

Für den genannten Ring $\{\mathrm{g}, \mathrm{u}\}$ und Variablen x und y gilt

$$(\mathrm{u}x + \mathrm{u}y)^2 = \mathrm{u}^2x^2 + \mathrm{u}^2xy + \mathrm{u}^2yx + \mathrm{u}^2y^2 = \mathrm{u}x^2 + \mathrm{g}xy + \mathrm{u}y^2.$$

Wenn man wie üblich das Nullelement g mit 0 und Einselement u mit 1 bezeichnet, wird es für Anfänger schon etwas verwirrend. Wenn man dann noch $1x$ durch x abkürzt und $0xy$ weglässt, so sieht man, dass in diesem Ring die ungewohnte binomischen Formel

$$(x + y)^2 = x^2 + y^2$$

gilt. Die beiden Seiten der Gleichung sind äquivalente Terme und ergeben somit beim Einsetzen von Ringelementen den selben Wert, wovon man sich auch direkt überzeugen kann.

Mit Termen aus Variablen und Elementen eines gegebenen Ringes R kann man mehrere Operationen ausführen: Man kann sie addieren, miteinander multiplizieren, und man kann eine Variable in einem Term durch einen anderen Term ersetzen (man sagt auch „substituieren“). Wenn man einen der beteiligten Terme durch einen äquivalenten ersetzt, dann geht in all diesen Fällen das Ergebnis in einen äquivalenten Term über. Folglich kann man diese Operationen auch mit Polynomen ausführen. Insbesondere ist die Menge der Polynome in[1] x und y mit Koeffizienten in einem gegebenen Ring R wieder ein Ring, den man mit $R[x, y]$ bezeichnet. Das Gleiche gilt natürlich auch für jede andere Menge von Variablen. Es ist üblich, ein Polynom durch einen Buchstaben abkürzen, hinter den man die vorkommenden Variablen in Klammern schreibt, z. B.

$$p(x, y) = (3x^2 + 2xy)(5x + y^2 + 4y).$$

Setzen wir hier $7x + z$ für y ein, so schreiben wir

$$p(x, 7x + z) = (3x^2 + 2x(7x + z))(5x + (7x + z)^2 + 4(7x + z)).$$

[1] Man sagt zwar „Funktion von den Variablen x und y“, aber „Polynom in den Variablen x und y“.

Ersetzt man alle vorkommenden Variablen durch Elemente von R, so erhält man ein Element von R, z. B.

$$p(2,1) = (3 \cdot 2^2 + 2 \cdot 2 \cdot 1)(5 \cdot 2 + 1^2 + 4 \cdot 1) = 16 \cdot 15 = 240.$$

Hier erhalten wir also eine Funktion, die z. B. dem geordneten Paar $(2,1)$ den Wert 240 zuordnet. Allgemein liefert ein Polynom in n Variablen mit Koeffizienten in einem Ring R also eine Funktion $p : R^n \to R$, für die man das selbe Formelzeichen benutzt. Es entsteht die Frage, ob verschiedene Polynome die selbe Funktion darstellen können. Zur Beantwortung benötigen wir einen weiteren Begriff.

Definition 3. *Ist R eine Teilmenge eines Ringes S, die das Null- und das Einselement von S enthält, abgeschlossen unter den Operationen von S ist und mit jedem Element auch sein entgegengesetztes Element enthält, dann nennen wir R einen Teilring oder Unterring von S.*

Diese Situation liegt im Fall der Zahlbereiche vor. So ist $\mathbf{Z}$ ein Unterring des Ringes $\mathbf{Q}$ und dieser wiederum ein Unterring von $\mathbf{R}$. Es ist klar, dass Terme mit Koeffizienten im Unterring R, die über R äquivalent sind, dann auch über dem gesamten Ring S äquivalent sein müssen. Jedes Polynom mit Koeffizienten in R können wir also auch als Polynom mit Koeffizienten in S ansehen, in das wir Elemente von S einsetzen können.

Aufgabe 4(b) zeigt uns, dass die Terme x und x^2 mit Koeffizienten im Ring $\{g, u\}$ beim Einsetzen von Elementen eines größeren Ringes manchmal verschiedene Werte annehmen. Also können sie auch über dem ursprünglichen Ring nicht äquivalent gewesen sein, obwohl sie dort nach Aufgabe 4(a) gleiche Werte annehmen. Das beantwortet unsere Frage mit „nein“. Es zeigt auch, dass es unklug wäre, Polynome als Funktionen zu definieren, weil dann ein Polynom mit Koeffizienten in einem Unterring nicht auf eindeutige Weise als Polynom mit Koeffizienten im gesamten Ring betrachtet werden könnte.

Um den Beweis von Satz 1 anzudeuten, brauchen wir noch einen Begriff.

Definition 4. *Es seien R und S Ringe. Eine Abbildung $h : R \to S$ wird Homomorphismus genannt, wenn sie folgende Eigenschaften hat.*

(i) Es gilt $h(0) = 0$ und $h(1) = 1$.

(ii) Für alle Elemente a und b von R gilt

$$h(a + b) = h(a) + h(b), \qquad h(a \cdot b) = h(a) \cdot h(b).$$

Man beachte, dass hier die Symbole 0, 1, + und − auf verschiedenen Seiten der Gleichung verschiedene Bedeutung haben: Links beziehen sie sich auf den Ring R und rechts auf S. Als Beispiel betrachten wir die Abbildung $h : \mathbf{Z} \to \{\mathrm{g}, \mathrm{u}\}$, die jeder geraden Zahl den Wert g und jeder ungeraden Zahl den Wert u zuordnet. Dann ist h ein Homomorphismus. Hier ist ein weiteres Beispiel: Für jeden Unterring R eines Ringes S ist die durch $i(a) = a$ gegebene Abbildung $i : R \to S$ ein Homomorphismus.

Lemma 1. *Es sei R ein Unterring des Ringes S und n eine natürliche Zahl. Des weiteren seien $x_1, \ldots, x_n$ Variable und $b_1, \ldots, b_n$ Elemente von S. Dann gibt es genau einen Homomorphismus $h : R[x_1, \ldots, x_n] \to S$, so dass $h(x_1) = b_1, \ldots, h(x_n) = b_n$ und $h(a) = a$ für alle $a \in R$.*

Zum Beweis des Lemmas nur soviel: Ist ein Polynom mit Koeffizienten in R gegeben, so können wir es nach dem Gesagten als Polynom mit Koeffizienten in S betrachten und für die Variablen $x_1, \ldots, x_n$ die Elemente $b_1, \ldots, b_n$ einsetzen. Der entstehende Wert in S ist dann das Bild des gegebenen Polynoms unter der Abbildung h.

Die meisten Autoren definieren einen anderen Begriff des Polynoms, den wir zur Unterscheidung Standardpolynom nennen wollen. Ein Standardpolynom in n Variablen mit Koeffizienten in einem Ring R ist eine Abbildung a, die jedem n-Tupel $(i_1, \ldots, i_n)$ von natürlichen Zahlen ein Element $a_{i_1,\ldots,i_n}$ von R zuordnet, wobei nur endlich viele Werte von Null verschieden sind. Für den Moment bezeichnen wir die Menge dieser Standardpolynome mit R_n. Es sei $g : R_n \to R[x_1, \ldots, x_n]$ die Abbildung, die einem Standardpolynom a das Polynom

$$\sum_{(i_1,\ldots,i_n)} a_{i_1,\ldots,i_n} x_1^{i_1} \ldots x_n^{i_n}$$

zuordnet. Es gibt nur eine Möglichkeit, eine Addition und eine Multiplikation von Standardpolynomen zu definieren, so dass R_n zu einem Ring und g zu einem Homomorphismus wird. Für diesen Ring beweist man ein Analogon von Lemma 1, und g ist der darin vorkommende Homomorphismus im Fall $S = R[x_1, \ldots, x_n]$ und $b_1 = x_1, \ldots, b_n = x_n$. Nun können wir das ursprüngliche Lemma 1 im Fall $S = R_n$ anwenden und erhalten einen Homomorphismus $h : R[x_1, \ldots, x_n] \to R_n$. Die Verkettung $g \circ h$ erfüllt dann ebenso wie die identische Abbildung die Bedingungen von Lemma 1 im Fall $S = R[x_1, \ldots, x_n]$, also muss wegen der Eindeutigkeitsaussage $g \circ h = \mathrm{id}$ sein. Ebenso beweist man, dass $h \circ g = \mathrm{id}$. Somit sind g und h zueinander inverse Homomorphismen und folglich die Ringe $R[x_1, \ldots, x_n]$ und R_n isomorph („von gleicher Gestalt“), so dass die Bezeichnung R_n letztlich überflüssig wird.

Aus der Bijektivität von g folgt, dass zwei Terme in Standardform genau dann äquivalent sind, wenn die Koeffizienten eines beliebigen Monoms in beiden übereinstimmen. Daraus folgen sofort die Aussagen von Satz 1.

3 Formen

Im Namen dieser Vorlesung steckt der Begriff „Form“. Das ist ein veraltetes Wort für „homogenes Polynom“.

Definition 5. *Es seien n und k natürliche Zahlen und $x_1, \ldots, x_n$ Variablen. Man sagt, ein Polynom $p(x_1, \ldots, x_n)$ mit Koeffizienten in einem Ring R sei homogen vom Grad k, wenn im Ring $R[x_1, \ldots, x_n, t]$ die Gleichheit*

$$p(tx_1, \ldots, tx_n) = t^k p(x_1, \ldots, x_n)$$

gilt, wobei t eine beliebige von $x_1, \ldots, x_n$ verschiedene Variable ist.

Bringen wir ein Polynom in Standardform, z. B.

$$p(x, y) = 15x^3 + 3x^2y^2 + 22x^2y + 2xy^3 + 8xy^2,$$

so erhalten wir sofort

$$p(tx, ty) = 15t^3x^3 + 3t^4x^2y^2 + 22t^3x^2y + 2t^4xy^3 + 8t^3xy^2.$$

Der Exponent von t ist in jedem Monom gleich der Summe der übrigen Exponenten. In der Standardform von $t^k p(x, y)$ hingegen hat t in allen Monomen den gleichen Exponenten k. Wir schließen mit Satz 1, dass ein Polynom genau dann homogen vom Grad k ist, wenn in allen vorkommenden Monomen die Summe der Exponenten gleich k ist. Das erklärt den Namen „homogen“. Natürlich ist ein Monom selbst ein homogenes Polynom, wobei sein Grad gleich der Summe der Exponenten ist.

Fassen wir in einem beliebigen Polynom Monome von gleichem Grad zusammen, so können wir das Polynom als Summe von homogenen Polynomen schreiben, die wir seine homogenen Komponenten nennen. Der Grad eines Polynoms ist der höchste Grad, der bei seinen Monomen vorkommt.

Ein homogenes Polynom vom Grad 1, 2 bzw. 3 nennt man auch heute noch Linearform, quadratische Form bzw. kubische Form. Beispiele sind

$$8x + 5y - 3z,$$
$$7x^2 + 2xy - 11xz - 4y^2 + 3yz + z^2,$$
$$2x^3 - 4x^2y + 3xz^2 - y^3 + 9y^2z + 5z^3.$$

Im Fall von n Variablen lässt sich eine Linearform so ausdrücken:

$$l(x_1, \ldots, x_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

Dabei stehen $a_1, \ldots, a_n$ für Elemente des Ringes, die durch die Linearform l festgelegt sind, während $x_1, \ldots, x_n$ unbestimmte Variablen sind. Nach dem Distributivgesetz hat eine Linearform die Eigenschaft

$$l(x_1 + y_1, \ldots, x_n + y_n) = l(x_1, \ldots, x_n) + l(y_1, \ldots, y_n). \tag{1}$$

Der Name „Linearform“ kommt daher, dass für eine nichtverschwindende Linearform $l(x, y)$ mit reellen Koeffizienten die Lösungsmenge der Gleichung $l(x, y) = 0$ eine gerade Line (kurz gesagt, eine Gerade) ist.

Definition 6. *Es seien m und n natürliche Zahlen und $x_1, \ldots, x_m, y_1, \ldots, y_n$ Variablen, die in zwei disjunkte Mengen aufgeteilt sind. Ein Polynom*

$$b(x_1, \ldots, x_m, y_1, \ldots, y_n)$$

mit Koeffizienten in einem Ring R heißt Bilinearform, wenn in dem Ring $R[x_1, \ldots, x_m, y_1, \ldots, y_n, t]$ die Gleichheiten

$$\begin{aligned} b(tx_1, \ldots, tx_m, y_1, \ldots, y_n) &= t\, b(x_1, \ldots, x_m, y_1, \ldots, y_n), \\ b(x_1, \ldots, x_m, ty_1, \ldots, ty_n) &= t\, b(x_1, \ldots, x_m, y_1, \ldots, y_n) \end{aligned}$$

gelten, wobei t eine beliebige von $x_1, \ldots, x_m, y_1, \ldots, y_n$ verschiedene Variable ist.

Angenommen, zwischen den Variablen der ersten und der zweiten Menge besteht eine umkherbar eindeutige Zuordnung (gegeben durch die Nummerierung, wobei natürlich $m = n$ sein muss). Die Bilinearform wird symmetrisch genannt, wenn

$$b(x_1, \ldots, x_n, y_1, \ldots, y_n) = b(y_1, \ldots, y_n, x_1, \ldots, x_n).$$

An der Standardform kann man leicht erkennen, dass ein Polynom genau dann eine Bilinearform ist, wenn in jedem Monom genau eine Variable aus der ersten Menge und eine aus der zweiten Menge jeweils in der ersten Potenz auftritt, z. B.

$$b(x_1, x_2, y_1, y_2, y_3) = 3x_1y_1 - 2x_1y_2 + 4x_1y_3 + x_2y_1 + 5x_2y_2 - x_2y_3.$$

Besteht eine umkehrbar eindeutige Zuordnung zwischen den Variablen der ersten und zweiten Menge, so entsteht – durch Ersetzung jeder Variablen durch die korrespondierende Variable – aus jedem Monom wieder ein Monom, z. B. aus dem Monom x_2y_3 das Monom y_2x_3. Eine Bilinearform ist genau dann symmetrisch, wenn je zwei in dieser Beziehung stehende Monome den gleichen Koeffizienten haben, z. B.

$$s(x_1, x_2, y_1, y_2) = 2x_1y_1 + 3x_1y_2 + 3x_2y_1 + x_2y_2.$$

Ist $m = n$, so erhalten wir aus einer Bilinearform $b(x_1, \ldots, x_n, y_1, \ldots, y_n)$ durch sogenannte Symmetrisierung eine symmetrische Bilinearform

$$s(x_1, \ldots, x_n, y_1, \ldots, y_n) = b(x_1, \ldots, x_n, y_1, \ldots, y_n) + b(y_1, \ldots, y_n, x_1, \ldots, x_n)$$

und durch Spezialisierung (nämlich Gleichsetzen korrespondierender Variablen) eine quadratische Form

$$q(x_1, \ldots, x_n) = b(x_1, \ldots, x_n, x_1, \ldots, x_n).$$

Umgekehrt gewinnen wir aus einer quadratischen Form $q(x_1, \ldots, x_n)$ durch sogenannte Polarisierung das Polynom

$$p(x_1, \ldots, x_n, y_1, \ldots, y_n) = q(x_1+y_1, \ldots, x_n+y_n) - q(x_1, \ldots, x_n) - q(y_1, \ldots, y_n).$$

Satz 2. *(i) Ist $b(x_1, \ldots, x_m, y_1, \ldots, y_n)$ eine Bilinearform, so gilt*

$$\begin{aligned} b(x_1 + u_1, \ldots, x_m + u_m, y_1, \ldots, y_n) &= b(x_1, \ldots, x_m, y_1, \ldots, y_n) \\ &\quad + b(u_1, \ldots, u_m, y_1, \ldots, y_n), \\ b(x_1, \ldots, x_m, y_1 + v_1, \ldots, y_n + v_n) &= b(x_1, \ldots, x_m, y_1, \ldots, y_n) \\ &\quad + b(x_1, \ldots, x_m, v_1, \ldots, v_n). \end{aligned}$$

(ii) Wenden wir auf eine Bilinearform erst die Spezialisierung und dann die Polarisierung an, so erhalten wir das Selbe wie durch Anwendung der Symmetrisierung.

(iii) Aus einer quadratischen Form entsteht durch Polarisierung eine symmetrische Bilinearform. Wenden wir darauf die Spezialisierung an, so erhalten wir das Doppelte der ursprünglichen quadratischen Form.

Beweis. Wir nehmen der Einfachheit halber an, dass $n = 2$ ist. Eine beliebige Bilinearform lässt sich dann so schreiben:

$$\begin{aligned} b(x_1, x_2, y_1, y_2) = \quad & a_{1,1}x_1y_1 + a_{1,2}x_1y_2 \\ & + a_{2,1}x_2y_1 + a_{2,2}x_2y_2. \end{aligned} \tag{2}$$

Dabei bezeichnen die Variablen $a_{i,j}$ feste Elemente des zugrunde liegenden Ringes. Die Aussagen (i) beweist man genauso wie die Eigenschaft (1) von Linearformen mit Hilfe des Distributivgesetzes. Wir lassen die Einzelheiten weg.

Nun folgt durch Anwendung der beiden Aussagen (i)

$$\begin{aligned} & b(x_1 + y_1, x_2 + y_2, x_1 + y_1, x_2 + y_2) \\ & \quad = b(x_1, x_2, x_1 + y_1, x_2 + y_2) + b(y_1, y_2, x_1 + y_1, x_2 + y_2) \\ & \quad = b(x_1, x_2, x_1, x_2) + b(x_1, x_2, y_1, y_2) + b(y_1, y_2, x_1, x_2) + b(y_1, y_2, y_1, y_2). \end{aligned}$$

Bezeichnen wir die Spezialisierung mit q und die Symmetrisierung mit s, so erhalten wir

$$q(x_1 + y_1, x_2 + y_2) = q(x_1, x_2) + s(x_1, x_2, y_1, y_2) + q(y_1, y_2).$$

Bringen wir alle Terme mit q auf die linke Seite, so erkennen wir dort die Polarisierung von q, und Aussage (ii) ist bewiesen.

Zum Beweis von Aussage (iii), wobei wir wieder $n = 2$ annehmen, betrachten wir eine beliebige quadratische Form, die wir wie folgt schreiben können:

$$q(x_1, x_2) = c_{1,1}x_1^2 + c_{1,2}x_1x_2 + c_{2,2}x_2^2. \tag{3}$$

Dabei sind die $c_{i,j}$ wieder Elemente des Ringes. Durch Polarisierung erhalten wir die Bilinearform

$$\begin{aligned} b(x_1, x_2, y_1, y_2) = c_{1,1}(x_1 + y_1)^2 + c_{1,2}(x_1 + y_1)(x_2 + y_2) + c_{2,2}(x_2 + y_2)^2 \\ -(c_{1,1}x_1^2 + c_{1,2}x_1x_2 + c_{2,2}x_2^2) \\ -(c_{1,1}y_1^2 + c_{1,2}y_1y_2 + c_{2,2}y_2^2). \end{aligned}$$

Nach dem Ausmultiplizieren kürzen sich alle Monome weg, in denen beide Variablen zur ersten oder beide zur zweiten Gruppe gehören, und es verbleibt

$$b(x_1, x_2, y_1, y_2) = 2c_{1,1}x_1y_1 + c_{1,2}x_1y_2 + c_{1,2}x_2y_1 + 2c_{2,2}x_2y_2,$$

wobei wir das Ringelement $1 + 1$ mit 2 abkürzen. Dies ist in der Tat eine symmetrische Bilinearform, und durch Spezialisierung erhalten wir $2q$.

Es ist leicht, diesen Beweis auf beliebiges n zu verallgemeinern, wobei die Formeln allerdings etwas unhandlich werden. □

Folgerung 1. *Ist das Element $1 + 1$ im Ring R invertierbar, so ist jede quadratische Form mit Koeffizienten in R die Spezialisierung einer symmetrischen Bilinearform.*

Dabei wird ein Element a von R invertierbar in R genannt, wenn es ein Element b von R mit der Eigenschaft $a \cdot b = 1$ gibt. Bezeichnen wir das Inverse von 2 mit $\frac{1}{2}$, so ist die in der Folgerung gemeinte Bilinearform das $\frac{1}{2}$-fache[2] der Polarisierung.

Wie wir wissen, definiert jedes Polynom in n Variablen mit Koeffizienten in einem Ring R eine Funktion $R^n \to R$, wobei im Allgemeinen verschiedene Polynome die selbe Funktion definieren können (vgl. Aufgabe 4). Aus den Darstellungen im obigen Beweis erhalten wir aber:

Lemma 2. *Eine Linearform, Bilinearform oder quadratische Form ist durch die zugehörige Funktion eindeutig bestimmt.*

[2]Viele Autoren nennen dies die Polarisierung, wofür sie die Invertierbarkeit von $1 + 1$ von Anbeginn voraussetzen.

Beweis. Wir zeigen, dass die Koeffizienten der Form durch die Funktionswerte eindeutig bestimmt sind. Setzen wir z. B. in Gleichung (1) für die Variable x_i den Wert 1 und für alle anderen Variablen den Wert 0 ein, so erhalten wir

$$l(0, \ldots, 0, 1, 0 \ldots, 0) = a_i.$$

Setzen wir in Gleichung (2) für x_i und y_j den Wert 1 und für alle anderen Variablen den Wert 0 ein, so erhalten wir $a_{i,j}$.

Bei einer quadratischen Form kann man die Bezeichnung $c_{j,i}$ als Synonym für $c_{i,j}$ betrachten, weil sich die Monome $x_i x_j$ und $x_j x_i$ nicht unterscheiden. Setzen wir in Gleichung (3) für x_i den Wert 1 und für alle anderen Variablen den Wert 0 ein, so erhalten wir $c_{i,i}$. Setzen wir hingegen für x_i und y_j den Wert 1 und für alle anderen Variablen den Wert 0 ein, wobei $i \neq j$, so erhalten wir $c_{i,i} + c_{i,j} + c_{j,j}$. Da $c_{i,i}$ und $c_{j,j}$ bereits eindeutig bestimmt sind, ist es auch $c_{i,j}$. □

Der Einfachheit halber lässt man die Kommas in den doppelten Indizes meist weg. Aus dem Zusammenhang wird klar, dass a_{12} nicht a mit dem Index zwölf bedeuten soll und a_{ij} nicht a mit dem Index $i \cdot j$.

4 Diagonalisierung

Manchmal lassen sich Polynome durch Substitutionen vereinfachen. Setzen wir etwa Linearformen in den Variablen $x_1, \ldots, x_n$ an Stelle der Variablen $u_1, \ldots, u_n$ in ein Polynom $q(u_1, \ldots, u_n)$ ein, so erhalten wir ein Polynom $p(x_1, \ldots, x_n)$, z. B.

$$q(x_1 + 3x_2 + 4x_3, x_1 + 4x_2 + 5x_3, x_1 + x_2 + x_3) = p(x_1, x_2, x_3).$$

Hat man das einfachere Polynom untersucht, möchte man die Substitution umkehren, um die Ergebnisse auf das ursprüngliche Polynom zu übertragen. Das ist in unserem Beispiel möglich. Es gilt nämlich, wie wir gleich begründen werden,

$$p(u_1 - u_2 + u_3, -4u_1 + 3u_2 + u_3, 3u_1 - 2u_2 - u_3) = q(u_1, \ldots, u_n).$$

Meist schreibt man einfach

$$p(x_1, \ldots, x_n) = q(u_1, \ldots, u_n),$$

wobei die Variablen durch die Substitution

$$\begin{aligned} u_1 &= x_1 + 3x_2 + 4x_3 \\ u_2 &= x_1 + 4x_2 + 5x_3 \\ u_3 &= x_1 + \ x_2 + \ x_3 \end{aligned}$$

zusammen hängen. Dann gilt in der Tat

$$\begin{aligned} u_1 - u_2 + u_3 &= (x_1 + 3x_2 + 4x_3) - (x_1 + 4x_2 + 5x_3) + (x_1 + x_2 + x_3) = x_1, \\ -4u_1 + 3u_2 + u_3 &= -4(x_1 + 3x_2 + 4x_3) + 3(x_1 + 4x_2 + 5x_3) + (x_1 + x_2 + x_3) = x_2, \\ 3u_1 - 2u_2 - u_3 &= 3(x_1 + 3x_2 + 4x_3) - 2(x_1 + 4x_2 + 5x_3) - (x_1 + x_2 + x_3) = x_3, \end{aligned}$$

Die Frage ist, wie man eine geeignete Substitution und ihre Umkehrung findet.

Strenggenommen sind die Variablen x_i und u_j nicht gleichberechtigt, denn die ersteren bezeichnen unabhängige Variablen, die letzteren hingegen bezeichnen Linearformen in den unabhängigen Variablen. Diesen Schönheitsfehler kann man aber beseitigen. Wir könnten von Anfang an Terme in den unabhängigen Variablen x_1, x_2, x_3, u_1, u_2, u_3 betrachten und die Äquivalenz von Termen neu definieren, indem wir neben den Rechengesetzen auch die drei ersten obigen Gleichungen zur Umformung zulassen. Die Äquivalenzklassen bilden dann einen Ring, der sowohl zum Ring der Polynome in x_1, x_2, x_3 als auch zum Ring der Polynome in u_1, u_2, u_3 isomorph ist und in dem die obigen sechs Gleichungen gelten. Nun sind auch die Rollen von p und q vertauschbar.

Wir betrachten hier den Fall einer quadratischen Form, z. B.

$$p(x_1, x_2, x_3) = x_1^2 + 6x_1x_2 - 2x_1x_3 + 7x_2^2 - 2x_2x_3.$$

Mit Hilfe der aus der Schule bekannten Methode der quadratischen Ergänzung können wir die ersten beiden Summanden $x_1^2 + 6x_1x_2$ zu einem vollständigen Quadrat ergänzen, nämlich

$$x_1^2 + 6x_1x_2 + 9x_2^2 = (x_1 + 3x_2)^2.$$

Subtrahieren wir hier auf beiden Seiten die Ergänzung $9x_2^2$, so erhalten wir einen Ausdruck, den wir für die ersten beiden Summanden in unsere quadratische Form einsetzen können. Das Ergebnis ist (nach Vereinfachung)

$$p(x_1, x_2, x_3) = (x_1 + 3x_2)^2 - 2x_1x_3 - 2x_2^2 - 2x_2x_3.$$

Das gemischte Monom, das die Variablen x_1 und x_2 enthielt, ist verschwunden. Statt dessen hätten wir auch das gemischte Monom mit x_1 und x_3 verschwinden lassen können, indem wir das Quadrat von $x_1 - x_3$ betrachten. Man kann aber auch in einem Schritt sämtliche gemischten Monome beseitigen, in denen x_1 vorkommt. Dazu beginnen wir von vorn und betrachten

$$(x_1 + 3x_2 - x_3)^2 = \underline{x_1^2 + 6x_1x_2 - 2x_1x_3} + 9x_2^2 - 6x_2x_3 + x_3^2.$$

Lösen wir diese Gleichung nach dem unterstrichenen Term auf und setzen das Ergebnis in unsere quadratische Form ein, so erhalten wir

$$p(x_1, x_2, x_3) = (x_1 + 3x_2 - x_3)^2 - 2x_2^2 + 4x_2x_3 - x_3^2.$$

Nun wenden wir die selbe Methode auf die gemischten Monome an, die x_2 enthalten. Den störenden Koeffizienten vor x_2^2 klammern wir zunächst aus:

$$-2x_2^2 + 4x_2x_3 = -2(x_2^2 - 2x_2x_3) = -2((x_2 - x_3)^2 - x_3^2) = -2(x_2 - x_3)^2 + 2x_3^2.$$

Einsetzen ergibt dann

$$p(x_1, x_2, x_3) = (x_1 + 3x_2 - x_3)^2 - 2(x_2 - x_3)^2 + x_3^2.$$

Damit haben wir die gesuchte lineare Substitution gefunden. Setzen wir nämlich

$$\begin{aligned} u_1 &= x_1 + 3x_2 - x_3, \\ u_2 &= \quad x_2 - x_3, \\ u_3 &= \quad\quad x_3, \end{aligned}$$

so ergibt sich p durch Einsetzen dieser Linearformen in die quadratische Form

$$q(u_1, u_2, u_3) = u_1^2 - 2u_2^2 + u_3^2,$$

die keine gemischten Monome hat.

Um unsere Substitutionen umzukehren, ersetzen wir gemäß der dritten Gleichung die Variable x_3 in den anderen Gleichungen durch u_3. Dann stellen wir die zweite Gleichung nach x_2 um und eliminieren mit Hilfe der gewonnenen Formel x_2 aus der ersten Gleichung. Schließlich stellen wir diese nach x_1 um. Hier sind die Ergebnisse zusammengefasst:

$$\begin{aligned} x_1 &= u_1 - 3u_2 - 2u_3, \\ x_2 &= \quad u_2 + \ u_3, \\ x_3 &= \quad\quad u_3. \end{aligned}$$

Es gibt aber ein Hindernis, wenn der Koeffizient von x_1^2 gleich Null ist. Auch nach der Beseitigung der gemischten Monome mit x_1 haben wir anscheinend ein Problem, wenn der entstehende Koeffizient von x_2^2 gleich Null ist. Solange aber das Quadrat einer späteren Variablen vorkommt, können wir uns dadurch aus der Affäre ziehen, dass wir die Rollen der Variablen vertauschen, etwa bei der quadratischen Form

$$x_1x_2 - x_2^2 + 2x_2x_3.$$

Scheinbar ausweglos ist die Situation, wenn nur gemischte Monome vorkommen. Dann hilft eine andere Methode: Die Substitution

$$\begin{aligned} x_1 &= v_1 + v_2, \\ x_2 &= v_1 - v_2 \end{aligned}$$

liefert nach der dritten binomischen Formel

$$x_1x_2 = v_1^2 - v_2^2,$$

also eine quadratische Form ohne gemischte Monome. Diese Substitution ist auch umkehrbar, denn indem wir die beiden Gleichungen addieren bzw. voneinander subtrahieren, erhalten wir

$$\begin{aligned} v_1 &= \frac{1}{2}x_1 + \frac{1}{2}x_2, \\ v_1 &= \frac{1}{2}x_1 - \frac{1}{2}x_2. \end{aligned}$$

Wenn jetzt noch gemischte Monome vorkommen, lässt sich unsere obige Methode anwenden. Mitunter sind also mehrere Substitutionen nötig, aber diese lassen sich zu einer einzigen verketten.

Dass all unsere Umformungen Polynome mit ganzzahligen Koeffizienten ergaben, ist ein zufälliger Umstand. Betrachten wir nämlich die quadratische Form

$$3x_1^2 - x_1x_2,$$

so liefert die Methode der quadratischen Ergänzung

$$3(x_1^2 - \frac{1}{3}x_1x_2) = 3\left(\left(x_1 - \frac{1}{6}x_2\right)^2 - \frac{1}{36}x_2^2\right) = 3\Big(x_1 - \frac{1}{6}x_2\Big)^2 - \frac{1}{12}x_2^2.$$

Um beliebige quadratische Formen mit Koeffizienten in einem Ring behandeln zu können, muss dieser eine Zusatzeigenschaft haben.

Definition 7. *Ein Ring wird Körper genannt, wenn jedes Element außer dem Nullelement invertierbar ist.*

Die Bereiche der rationalen Zahlen und der reellen Zahlen sind Körper, der Bereich der ganzen Zahlen nicht. Weitere Beispiele für Körper sind die Ringe aus Aufgabe 3 und Präsenzaufgabe 2.

Natürlich überträgt sich die beschriebene Methode auf quadratische Formen in einer beliebigen Anzahl von Variablen mit Koeffizienten in einem Körper. Damit haben wir Folgendes bewiesen.

Satz 3. *Es sei*

$$\begin{aligned} p(x_1, \ldots, x_n) = c_{11}x_1^2 + c_{12}x_1x_2 + \ldots + c_{1n}x_1x_n \\ + c_{22}x_2^2 \quad + \ldots + c_{2n}x_2x_n \\ \ddots \qquad \vdots \\ + c_{nn}x_n^2 \end{aligned}$$

eine quadratische Form mit Koeffizienten c_{ij} in einem Körper K, in dem $1+1 \neq 0$ ist. Dann gibt es eine umkehrbare lineare Substitution

$$\begin{aligned} u_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n, \\ u_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n, \\ &\vdots \\ u_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{aligned} \tag{4}$$

mit Koeffizienten $a_{ij} \in K$, die p in eine quadratische Form der Art

$$q(u_1, \ldots, u_n) = d_1u_1^2 + d_2u_2^2 + \cdots + d_nu_n^2 \tag{5}$$

mit Koeffizienten $d_i \in K$ überführt.

Ordnet man die Terme von q in einem quadratischen Schema an wie die von p, so stehen sie alle auf der Diagonalen. Daher spricht man von einem Diagonalisierungsverfahren.

Betrachtet man quadratische Formen und Substitutionen mit *reellen* Koeffizienten, so lässt sich eine diagonalisierte Form noch weiter vereinfachen. Wir hatten oben zum Beispiel die Form

$$u_1^2 - 2u_2^2 + u_3^2$$

erhalten. Wegen $2u_2^2 = (\sqrt{2}u_2)^2$ erhalten wir vermittels der Substitution

$$\begin{aligned} v_1 &= u_1 \\ v_2 &= \sqrt{2}u_2 \\ v_3 &= u_3 \end{aligned}$$

die quadratische Form

$$v_1^2 - v_2^2 + v_3^2.$$

Man kann die v_j natürlich auch direkt durch die x_i ausdrücken und umgekehrt. Ganz allgemein kann man durch umkehrbare lineare Substitutionen mit reellen Koeffizienten eine diagonale Form gewinnen, in der nur die Koeffizienten 0, 1 und -1 vorkommen.

5 Moduln und Vektorräume

Wir hatten gesehen, dass ein Polynom $p(x_1, \ldots, x_n)$ mit Koeffizienten in einem Ring R eine Funktion von n Variablen definiert. Ihr Definitionsbereich

ist die Menge R^n aller n-Tupel $(x_1, \ldots, x_n)$ mit Einträgen $x_i \in R$. Häufig kürzt man ein solches n-Tupel auch durch einen einzelnen Buchstaben $\mathbf{x}$ ab, den wir durch Fettdruck hervorheben.[3] Man definiert eine Addition von n-Tupeln und eine Multiplikation mit Elementen a von R, genannt Skalarmultiplikation, wie folgt:

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$
$$a \cdot (x_1, \ldots, x_n) = (a \cdot x_1, \ldots, a \cdot x_n).$$

Diese Operationen haben gewisse Eigenschaften. Sie treten auch bei anderen Objekten auf, die man als Moduln bezeichnet.

Definition 8. *Auf einer Menge V sei eine Operation $+$ gegeben, für die das Kommutativgesetz und das Assoziativgesetz gelten, wobei es ein Nullelement $\mathbf{0}$ gibt und zu jedem Element von V genau ein entgegengesetztes Element.*

Außerdem sei ein Ring R mit seinen Operationen $+$ und $\cdot$ gegeben sowie eine weitere Operation, die einem Element $a \in R$ und einem Element $\mathbf{x} \in V$ ein Element $a \cdot \mathbf{x} \in V$ zuordnet, wobei für alle a, $b \in R$ und $\mathbf{x}$, $\mathbf{y} \in V$ gilt

$$a \cdot (\mathbf{x} + \mathbf{y}) = a \cdot \mathbf{x} + a \cdot \mathbf{y}, \qquad (a + b) \cdot \mathbf{x} = a \cdot \mathbf{x} + b \cdot \mathbf{x},$$
$$a \cdot (b \cdot \mathbf{x}) = (a \cdot b) \cdot \mathbf{x}, \qquad 1 \cdot \mathbf{x} = \mathbf{x}.$$

In diesem Fall nennt man V einen Modul über dem Ring R. Ist R ein Körper, so nennt man V einen Vektorraum, und seine Elemente nennt man Vektoren.

Insbesondere ist die Menge R^n mit den oben definierten Operationen ein Modul. Es gibt aber auch andere Beispiele.

So kann man die Menge der Verschiebungen einer Ebene als Vektorraum über dem Körper $\mathbf{R}$ der reellen Zahlen auffassen. Dabei wird die Addition definiert als Nacheinanderausführung von Verschiebungen und die Skalarmultiplikation als Streckung der Verschiebungslänge. Wählt man zwei nichtparallele Vektoren $\mathbf{v}_1$ und $\mathbf{v}_2$, so kann man auf intuitivem Niveau begründen, dass sich jeder Vektor $\mathbf{x}$ in der Form $\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2$ schreiben lässt. Die Zahlen x_1 und x_2 nennt man dann die Koordinaten von x bezüglich der Basis $\mathbf{v}_1$, $\mathbf{v}_2$. Auf diese Weise entspricht jedem Vektor ein geordnetes Paar $(x_1, x_2) \in \mathbf{R}^2$ und umgekehrt. Wir erhalten also eine umkehrbar eindeutige Zuordnung zwischen dem Vektorraum der Verschiebungen und dem Vektorraum $\mathbf{R}^2$.

Ähnlich beschreibt man Verschiebungen des Raumes durch Tripel reeller Zahlen (x_1, x_2, x_3). Betrachtet man nur diejenigen Verschiebungen, die ein gegebenes Kristallgitter in sich überführen, so erhält man einen Modul über dem Ring $\mathbf{Z}$ der ganzen Zahlen.

[3] Auch die Schreibweise $\vec{x}$ ist üblich.

Definition 9. *Eine Folge von Elementen* $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *eines Moduls* V *über einem Ring* R *heißt Basis, wenn sich jedes Element* $\mathbf{x}$ *von* V *in der Form*

$$\mathbf{x} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n \tag{6}$$

mit eindeutig bestimmten Elementen $x_1, \ldots, x_n$ *von* R *schreiben lässt. Ein freier Modul vom Rang* n *ist ein Modul, der eine Basis aus* n *Vektoren besitzt.*

Es gibt tatsächlich Moduln, die unendliche Basen oder überhaupt keine Basen besitzen, aber die werden uns hier nicht interessieren. Den Ausdruck auf der rechten Seite der Gleichung (6) nennt man übrigens eine Linearkombination der Elemente $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Ist $\mathbf{w}_1, \ldots, \mathbf{w}_n$ eine weitere Basis[4] des Moduls V, so kann man das selbe Element $\mathbf{x}$ auch in dieser Basis ausdrücken, nämlich

$$\mathbf{x} = u_1\mathbf{w}_1 + \cdots + u_n\mathbf{w}_n$$

mit Koordinaten $u_j \in R$. Wie lassen sich diese aus den Koordinaten x_i bezüglich der ursprünglichen Basis errechnen? Nach Definition lässt sich auch jedes Element der alten Basis durch die neue Basis ausdrücken, also

$$\begin{aligned}
\mathbf{v}_1 &= a_{11}\mathbf{w}_1 + \cdots + a_{n1}\mathbf{w}_n \\
\mathbf{v}_2 &= a_{12}\mathbf{w}_1 + \cdots + a_{n2}\mathbf{w}_n \\
&\vdots \\
\mathbf{v}_n &= a_{1n}\mathbf{w}_1 + \cdots + a_{nn}\mathbf{w}_n
\end{aligned}$$

mit gewissen Elementen $a_{ij} \in R$. Setzen wir dies ein, so erhalten wir

$$\begin{aligned}
\mathbf{x} = {} & x_1(a_{11}\mathbf{w}_1 + \cdots + a_{n1}\mathbf{w}_n) \\
& + x_2(a_{12}\mathbf{w}_1 + \cdots + a_{n2}\mathbf{w}_n) \\
& \vdots \\
& + x_n(a_{1n}\mathbf{w}_1 + \cdots + a_{nn}\mathbf{w}_n)
\end{aligned}$$

und nach Ausmultiplizieren und Ausklammern der $\mathbf{w}_j$

$$\begin{aligned}
\mathbf{x} = {} & (a_{11}x_1 + \cdots + a_{1n}x_n)\mathbf{w}_1 \\
& + (a_{21}x_1 + \cdots + a_{2n}x_n)\mathbf{w}_2 \\
& \vdots \\
& + (a_{n1}x_1 + \cdots + a_{nn}x_n)\mathbf{w}_n
\end{aligned}$$

[4] Für viele Ringe (insbesondere für alle Körper) kann man zeigen, dass alle Basen eines Moduls die gleiche Anzahl von Elementen haben.

Da die Koordinaten bezüglich einer Basis eindeutig bestimmt sind, erhalten wir schließlich die selben Formeln wie in Gleichung (4) in Satz 3. Der Übergang zu einer anderen Basis bedeutet für die Koordinaten eines Elementes also eine lineare Substitution. Da wir die Rollen der beiden Basen vertauschen können, ist diese Substitution umkehrbar.
Beispiel. In einem Modul V über dem Ring $\mathbf{Z}$ mit der Basis $\mathbf{v}_1$, $\mathbf{v}_2$ wählen wir die Elemente

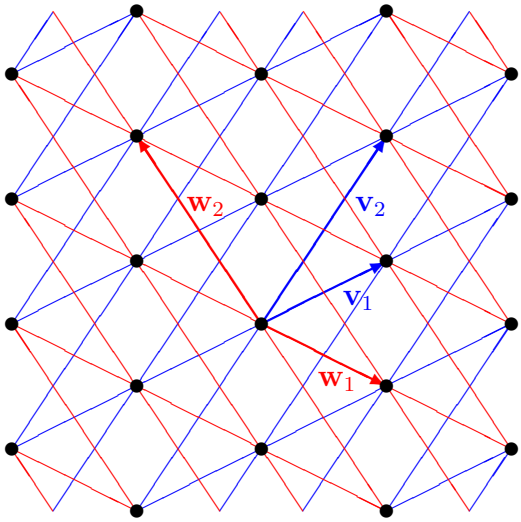
$$\begin{aligned} \mathbf{w}_1 &= 2\mathbf{v}_1 - \mathbf{v}_2 \\ \mathbf{w}_2 &= -3\mathbf{v}_1 + 2\mathbf{v}_2 \end{aligned}$$

Wir können diese Gleichungen nach $\mathbf{v}_1$ und $\mathbf{v}_2$ auflösen: Aus der ersten Gleichung erhalten wir $\mathbf{v}_2 = 2\mathbf{v}_1 - \mathbf{w}_1$. Damit können wir $\mathbf{v}_2$ aus der zweiten Gleichung eliminieren:

$$\mathbf{w}_2 = -3\mathbf{v}_1 + 2(2\mathbf{v}_1 - \mathbf{w}_1) = \mathbf{v}_1 - 2\mathbf{w}_1.$$

Lösen wir dies nach $\mathbf{v}_1$ auf und setzen das Ergebnis in die umgestellte erste Gleichung ein, so ergibt sich

$$\begin{aligned} \mathbf{v}_1 &= 2\mathbf{w}_1 + \mathbf{w}_2 \\ \mathbf{v}_2 &= 3\mathbf{w}_1 + 2\mathbf{w}_2 \end{aligned}$$



Ist nun $\mathbf{x} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2$, so folgt nach den obigen allgemeinen Berechnungen, dass $\mathbf{x} = u_1\mathbf{w}_1 + u_2\mathbf{w}_2$, wobei

$$\begin{aligned} u_1 &= 2x_1 + 3x_2 \\ u_2 &= x_1 + 2x_2 \end{aligned}$$

Ist umgekehrt eine Linearkombination $\mathbf{x} = u_1\mathbf{w}_1 + u_2\mathbf{w}_2$ gegeben, so berechnet man ihre Koordinaten bezüglich der Basis $\mathbf{v}_1$, $\mathbf{v}_2$ zu

$$\begin{aligned} x_1 &= \ 2u_1 - 3u_2 \\ x_2 &= -u_1 + 2u_2 \end{aligned}$$

Die Zahlen u_1 und u_2 sind als Lösung dieses Gleichungssystems eindeutig durch x_1, x_2, also durch $\mathbf{x}$ bestimmt, und somit ist auch $\mathbf{w}_1$, $\mathbf{w}_2$ eine Basis von V.

6 Formen auf Moduln

Wir definieren nun die Begriffe Linearform und Bilinearform auf scheinbar neue Weise.

Definition 10. *Es seien V und W freie Moduln über einem Ring R. Eine Abbildung $l : V \to R$ heißt Linearform, wenn für alle $\mathbf{x}$, $\mathbf{v} \in V$ und alle $t \in R$ gilt*

$$l(\mathbf{x} + \mathbf{v}) = l(\mathbf{x}) + l(\mathbf{v}), \qquad l(t \cdot \mathbf{x}) = t \cdot l(\mathbf{x}).$$

Eine Abbildung $b : V \times W \to R$ heißt Bilinearform, wenn für alle $\mathbf{x}$, $\mathbf{v} \in V$, alle $\mathbf{y}$, $\mathbf{w} \in W$ und alle $t \in R$ gilt

$$\begin{aligned} b(\mathbf{x} + \mathbf{v}, \mathbf{y}) &= b(\mathbf{x}, \mathbf{y}) + b(\mathbf{v}, \mathbf{y}), & b(\mathbf{x}, \mathbf{y} + \mathbf{w}) &= b(\mathbf{x}, \mathbf{y}) + b(\mathbf{x}, \mathbf{w}), \\ b(t \cdot \mathbf{x}, \mathbf{y}) &= t \cdot b(\mathbf{x}, \mathbf{y}), & b(\mathbf{x}, t \cdot \mathbf{y}) &= t \cdot b(\mathbf{x}, \mathbf{y}). \end{aligned}$$

Kennen wir die Werte einer Linearform l auf allen Elementen einer Basis von V, sagen wir

$$l(\mathbf{v}_1) = c_1, \quad \ldots, \quad l(\mathbf{v}_n) = c_n,$$

so können wir den Wert auf einem Element $\mathbf{x} \in V$ aus seinen Koordinaten $x_1, \ldots, x_n$ bezüglich dieser Basis errechnen. Nach Definition ist nämlich

$$\begin{aligned} l(x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n) &= l(x_1\mathbf{v}_1) + \cdots + l(x_n\mathbf{v}_n) \\ &= x_1 l(\mathbf{v}_1) + \cdots + x_n l(\mathbf{v}_n), \end{aligned}$$

so dass

$$l(\mathbf{x}) = c_1x_1 + \cdots + c_nx_n.$$

Der Wert einer Linearform im neuen Sinne auf einem Element $\mathbf{x}$ ist also gleich einer Linearform im alten Sinne, ausgewertet auf den Koordinaten von $\mathbf{x}$.

Ähnliches gilt für Bilinearformen. Ist $\mathbf{v}_1, \ldots, \mathbf{v}_m$ eine Basis von V und $\mathbf{w}_1, \ldots, \mathbf{w}_n$ eine Basis von W, so errechnet sich der Wert einer Bilinearform auf Elementen $\mathbf{x} \in V$ und $\mathbf{y} \in W$ aus den Werten

$$b(\mathbf{v}_i, \mathbf{w}_j) = g_{ij}$$

nach der Formel

$$\begin{aligned} b(\mathbf{x}, \mathbf{y}) = {} & g_{11}x_1y_1 + g_{12}x_1y_2 + \cdots + g_{1n}x_1y_n \\ & + g_{21}x_2y_1 + g_{22}x_2y_2 + \cdots + g_{2n}x_2y_n \\ & \vdots \\ & + g_{m1}x_my_1 + g_{m2}x_my_2 + \cdots + g_{mn}x_my_n. \end{aligned}$$

Der Wert einer Bilinearform im neuen Sinne auf Elementen $\mathbf{x}$ und $\mathbf{y}$ ist also gleich einer Bilinearform im alten Sinne, ausgewertet auf den Koordinaten von $\mathbf{x}$ und $\mathbf{y}$.

Nun betrachten wir den Fall, dass W der selbe Modul wie V ist. Dann können wir die Symmetrisierung einer Bilinearform $b : V \times V \to R$ durch

$$s(\mathbf{x}, \mathbf{y}) = b(\mathbf{x}, \mathbf{y}) + b(\mathbf{y}, \mathbf{x})$$

und die Spezialisierung von b durch

$$q(\mathbf{x}) = b(\mathbf{x}, \mathbf{x})$$

definieren.

Definition 11. *Eine quadratische Form auf einem Modul V über einem Ring R ist eine Funktion $q : V \to R$, die durch Spezialisierung aus einer Bilinearform $b : V \times V \to R$ entsteht.*

Definieren wir die Polarisierung einer quadratischen Form durch

$$p(\mathbf{x}, \mathbf{y}) = q(\mathbf{x} + \mathbf{y}) - q(\mathbf{x}) - q(\mathbf{y}),$$

so gilt Satz 2 sinngemäß. Die dort vorkommende umkehrbbar eindeutige Zuordnung zwischen den Variablen, welche die Koordinaten von $\mathbf{x}$ und von $\mathbf{y}$ beschreiben, entsteht dadurch, dass wir gleiche Basen wählen, also $\mathbf{w}_1 = \mathbf{v}_1$, $\ldots$, $\mathbf{w}_n = \mathbf{v}_n$.

Die Spezialisierung q einer symmetrischen Bilinearform b wird in einer Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ durch eine quadratische Form in n Variablen beschrieben. Gehen wir zu einer anderen Basis $\mathbf{w}_1, \ldots, \mathbf{w}_n$ über, so ändert sich diese quadratische Form mittels einer linearen Substitution, wie im vorigen Abschnitt

beschrieben. Dabei wird die quadratische Form genau dann diagonalisiert, wenn

$$b(\mathbf{w}_i, \mathbf{w}_j) = 0 \qquad \text{für } i \neq j.$$

Eine Basis mit dieser Eigenschaft nennt man eine *Orthogonalbasis* für die Form b. Der in Satz 3 vorkommende Koeffizient von u_i^2 ist dann

$$d_i = q(\mathbf{w}_i).$$

Definition 12. *Es sei b eine symmetrische Bilinearform b auf einem Vektorraum V über einem Körper K und q ihre Spezialisierung.*

- *Man sagt, b sei ausgeartet, wenn es ein von Null verschiedenes Element $\mathbf{x} \in V$ gibt, so dass für alle $\mathbf{y} \in V$ gilt $b(\mathbf{x}, \mathbf{y}) = 0$.*

Nun sei K der Körper $\mathbf{R}$ der reellen Zahlen und die Bilinearform b nicht ausgeartet.

- *Man sagt, b sei definit, wenn die Werte $q(\mathbf{x})$ für alle $\mathbf{x} \neq 0$ das selbe Vorzeichen haben. Andernfalls nennt man b indefinit.*

Bei definiten Bilinearformen kann man offensichtlich zwischen positiv definiten und negativ definiten unterscheiden.

Satz 4. *Es sei b eine symmetrische Bilinearform b auf einem Vektorraum V über einem Körper K und q ihre Spezialisierung. Weiter sei $d_i = b(\mathbf{v}_i, \mathbf{v}_i)$, wobei $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eine Orthogonalbasis für b ist.*

(i) Die Form b ist genau dann ausgeartet, wenn unter den Elementen $d_1, \ldots, d_n$ das Nullelement vorkommt.

Nun sei K der Körper $\mathbf{R}$ der reellen Zahlen und die Bilinearform b nicht ausgeartet.

(ii) Die Form b ist genau dann positiv definit, wenn $d_1 > 0, \ldots, d_n > 0$.

(iii) Die Form b ist genau dann negativ definit, wenn $d_1 < 0, \ldots, d_n < 0$.

Beweis. **(i)** Ist der Koeffizient d_i gleich Null, so gilt $b(\mathbf{v}_i, \mathbf{y}) = 0$ für alle $\mathbf{y} \in V$, also ist b ausgeartet. Ist umgekehrt b ausgeartet, so gibt es einen von Null verschiedenen Vektor $\mathbf{x} \in V$, so dass für alle i gilt $d_i x_i = b(\mathbf{x}, \mathbf{v}_i) = 0$, wobei x_i die $i$te Koordinate von $\mathbf{x}$ bezeichnet. Wegen $\mathbf{x} \neq \mathbf{0}$ gibt es ein i mit der Eigenschaft $x_i \neq 0$, und dann ist $d_i = 0$.

(ii) Ist b positiv definit, so gilt $d_i = q(\mathbf{v}_i) > 0$ für alle i. Sind umgekehrt alle d_i positiv, so sind in Gleichung (5) alle Terme nichtnegativ. Ist außerdem $\mathbf{x} \neq \mathbf{0}$, so ist wenigstens eine Koordinate x_i nicht Null. Also ist wenigstens ein Term positiv und somit $q(\mathbf{x}) > 0$. Analog beweist man (iii). □

Wie schon erwähnt, läuft die Diagonalisierung einer quadratischen Form darauf hinaus, eine Orthogonalbasis für die zugehörige symmetrische Bilinearform zu finden. Dieses Verfahren lässt sich auch in der Sprache von Basen formulieren und wird *Gram-Schmidtsches*[5] *Orthogonalisierungsverfahren* genannt. Dabei betrachtet man gewöhnlich eine positiv definite Bilinearform b auf einem reellen Vektorraum V und eine gegebene Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ von V. Man setzt $\mathbf{w}_1 = \mathbf{v}_1$ und sucht einen Vektor der Form

$$\mathbf{w}_2 = \mathbf{v}_2 + a_{21}\mathbf{w}_1,$$

der zu $\mathbf{w}_1$ bezüglich b orthogonal ist, d. h.

$$b(\mathbf{v}_2, \mathbf{w}_1) + a_{21}b(\mathbf{w}_1, \mathbf{w}_1) = b(\mathbf{w}_2, \mathbf{w}_1) = 0,$$

und weil $b(\mathbf{w}_1, \mathbf{w}_1) > 0$ ist,

$$a_{21} = -\frac{b(\mathbf{v}_2, \mathbf{w}_1)}{b(\mathbf{w}_1, \mathbf{w}_1)}.$$

Durch Einsetzen erhält man einen eindeutig bestimmten Vektor $\mathbf{w}_2$. Dieser kann nicht Null sein, weil sich sonst $\mathbf{v}_2$ durch $\mathbf{v}_1$ ausdrücken ließe, was bei einer Basis unmöglich ist. Somit ist auch $b(\mathbf{w}_2, \mathbf{w}_2) > 0$.

Als Nächstes sucht man einen Vektor der Form

$$\mathbf{w}_3 = \mathbf{v}_3 + a_{32}\mathbf{w}_2 + a_{31}\mathbf{w}_1,$$

der zu $\mathbf{w}_1$ und zu $\mathbf{w}_2$ orthogonal ist, d. h.

$$\begin{aligned} b(\mathbf{v}_3, \mathbf{w}_2) + a_{32}b(\mathbf{w}_2, \mathbf{w}_2) + a_{31}b(\mathbf{w}_1, \mathbf{w}_2) &= b(\mathbf{w}_3, \mathbf{w}_1) = 0, \\ b(\mathbf{v}_3, \mathbf{w}_1) + a_{32}b(\mathbf{w}_2, \mathbf{w}_1) + a_{31}b(\mathbf{w}_1, \mathbf{w}_1) &= b(\mathbf{w}_3, \mathbf{w}_2) = 0, \end{aligned}$$

und wegen $b(\mathbf{w}_1, \mathbf{w}_2) = b(\mathbf{w}_2, \mathbf{w}_1) = 0$ erhalten wir

$$a_{32} = -\frac{b(\mathbf{v}_3, \mathbf{w}_2)}{b(\mathbf{w}_2, \mathbf{w}_2)}, \qquad a_{31} = -\frac{b(\mathbf{v}_3, \mathbf{w}_1)}{b(\mathbf{w}_1, \mathbf{w}_1)}.$$

Wie oben sieht man, dass $\mathbf{w}_3 \neq 0$ ist, weil sich sonst $\mathbf{v}_3$ durch $\mathbf{v}_2$ und $\mathbf{v}_1$ ausdrücken ließe, und somit ist $b(\mathbf{w}_3, \mathbf{w}_3) > 0$.

Hat man bereits $\mathbf{w}_1, \ldots, \mathbf{w}_{k-1}$ bestimmt, so sieht man analog, dass man

$$\mathbf{w}_k = \mathbf{v}_k - \frac{b(\mathbf{v}_k, \mathbf{w}_{k-1})}{b(\mathbf{w}_{k-1}, \mathbf{w}_{k-1})}\mathbf{w}_{k-1} - \ldots - \frac{b(\mathbf{v}_k, \mathbf{w}_1)}{b(\mathbf{w}_1, \mathbf{w}_1)}\mathbf{w}_1$$

[5] nach Jørgen Pedersen Gram (1850-1916), der es in der Wahrscheinlichkeitstheorie benutzte, und Erhard Schmidt (1876-1959), der es auf unendlichdimensionale Räume verallgemeinerte.

setzen muss, solange $k < n$ ist. Am Ende hat man eine Orthogonalbasis $\mathbf{w}_1$, $\ldots$, $\mathbf{w}_n$ gefunden, denn man kann die Vektoren $\mathbf{v}_1$, $\ldots$, $\mathbf{v}_n$ durch $\mathbf{w}_1$, $\ldots$, $\mathbf{w}_n$ ausdrücken, also einen beliebigen Vektor $\mathbf{x} \in V$ in der Form

$$\mathbf{x} = u_1\mathbf{w}_1 + \ldots + u_n\mathbf{w}_n$$

darstellen, und wegen

$$b(\mathbf{x}, \mathbf{w}_i) = u_i b(\mathbf{w}_i, \mathbf{w}_i)$$

sind die Elemente

$$u_i = \frac{b(\mathbf{x}, \mathbf{w}_i)}{b(\mathbf{w}_i, \mathbf{w}_i)}$$

eindeutig bestimmt.

7 Kurven und Flächen zweiter Ordnung

Eine algebraische Gleichung ist eine Gleichung zwischen Termen, in denen Variablen x_1, $\ldots$, x_n mit Elementen eines Ringes durch Addition und Multiplikation verknüpft sind. Da man alle Terme auf eine Seite bringen kann, hat eine solche Gleichung die Form

$$p(x_1, \ldots, x_n) = 0,$$

wobei p ein Polynom ist. Man interessiert sich für die Lösungsmengen von algebraischen Gleichungen.

Ist der Ring ein Körper, so könnte man Terme zulassen, in denen die Division vorkommt. Man kann aber jeden solchen Term als Quotienten zweier Polynome p_1 und p_2 schreiben, und die Lösungsmenge der Gleichung

$$\frac{p_1(x_1, \ldots, x_n)}{p_2(x_1, \ldots, x_n)} = 0$$

ist die Differenzmenge aus den Lösungsmengen der Gleichungen

$$p_1(x_1, \ldots, x_n) = 0 \qquad \text{und} \qquad p_2(x_1, \ldots, x_n) = 0.$$

In der Schule werden Gleichungen ersten und zweiten Grades in einer Variablen behandelt, wobei man letztere durch die Methode der quadratischen Ergänzung einer Lösung zuführt. Wir wollen nun Gleichungen zweiten Grades von mehreren Variablen betrachten. Es sei also ein Polynom $p(x_1, \ldots, x_n)$ zweiten Grades mit Koeffizienten in einem Körper K gegeben. Dieses können wir in seine homogenen Komponenten zerlegen:

$$p(x_1, \ldots, x_n) = q(x_1, \ldots, x_n) + l(x_1, \ldots, x_n) + c,$$

wobei q eine quadratische Form, l eine Linearform und c eine Konstante ist. Wir nehmen nun an, dass in K gilt $1 + 1 \neq 0$. Dann können wir nach Satz 3 die quadratische Form q durch eine umkehrbare lineare Substitution diagonalisieren, d. h.

$$q(x_1, \ldots, x_n) = d_1 u_1^2 + \ldots + d_n u_n^2.$$

Nehmen wir die selbe Substitution in der Linearform l vor, so erhalten wir wieder eine Linearform, also

$$l(x_1, \ldots, x_n) = a_1 u_1 + \ldots + a_n u_n.$$

Betrachten wir nun die Terme mit einer Variablen u_i. Ist der Koeffizient $d_i \neq 0$, so hat $2d_i$ ein Inverses b_i, und wir finden die quadratische Ergänzung:

$$d_i u_i^2 + a_i u_i = d_i(u_i^2 + 2a_i b_i u_i) = d_i(u_i + a_i b_i)^2 - d_i a_i^2 b_i^2.$$

Mit der umkehrbaren Substitution

$$w_i = u_i + a_i b_i$$

wird dies zu $w_i^2 - d_i a_i^2 b_i^2$. Ist hingegen $d_i = 0$, aber $a_i \neq 0$, so ist die Substitution

$$w_i = -a_i u_i$$

umkehrbar. Im Fall $d_i = a_i = 0$ setzen wir einfach $w_i = u_i$.

Wir bezeichnen die Anzahl der Indizes i, für die $d_i \neq 0$ ist, mit k. Des Weiteren bezeichnen die Anzahl der Indizes i, für die $d_i \neq 0$ oder $a_i \neq 0$ ist, mit m. Wir können durch Umnummerierung erreichen, dass die Koeffizienten d_i für $i \leq k$ von Null verschieden und für $i > k$ gleich Null sind, und dass die a_i für $k < i \leq m$ von Null verschieden und für $i > m$ gleich Null sind. Fassen wir die Konstanten zusammen, so erhalten wir

$$p(x_1, \ldots, x_n) = d_1 w_1^2 + \ldots + d_k w_k^2 - w_{k+1} - \ldots - w_m - e$$

mit $e \in K$. Ist $m > k$, so kann man die Definition von w_{k+1} abändern:

$$w_{k+1} = -a_{k+1} u_{k+1} - \ldots - a_m u_m - e.$$

Dann erhalten wir

$$p(x_1, \ldots, x_n) = d_1 w_1^2 + \ldots + d_k w_k^2 - w_{k+1}.$$

Ist K der Körper der reellen Zahlen, so können wir die Definition von w_i für $i \leq k$ so abändern, dass an Stelle von d_i hier 1 oder -1 steht. Kommen keine

linearen Monome vor, also im Fall $k = m$, so ist ein n-Tupel $(w_1, \ldots, w_n)$ genau dann eine Lösung, wenn $(-w_1, \ldots, -w_n)$ eine Lösung ist.

Um Informationen über die Lösungsmenge der vereinfachten Gleichung

$$d_1 w_1^2 + \ldots + d_k w_k^2 = e \qquad \text{oder} \qquad d_1 w_1^2 + \ldots + d_k w_k^2 = w_{k+1}$$

in Informationen über die Lösungsmenge der ursprünglichen Gleichung zu übersetzen, müssen wir die vorgenommenen Substitutionen miteinander verketten. Man sieht, dass sich die Variablen $w_1, \ldots, w_n$ durch Polynome ersten Grades in den Variablen $x_1, \ldots, x_n$ ausdrücken lassen und umgekehrt. Dies nennt man eine umkehrbare affine Substitution. Der Grad eines Polynoms ändert sich bei einer affinen Substitution nicht.

Wir wollen nun die erhaltenen Erkenntnisse geometrisch deuten, wobei zunächst $n = 2$ sei. Dazu betrachten wir eine Ebene und interpretieren die Menge ihrer Verschiebungen als reellen Vektorraum V. Sind P und Q Punkte der Ebene, so gibt es genau eine Verschiebung, die P auf Q abbildet. Wir bezeichnen sie mit $\overrightarrow{PQ}$.

Nun halten wir in der Ebene einen Punkt O fest. Dann kann man jedem Vektor $\mathbf{x} \in V$ einen Punkt P der Ebene zuordnen, nämlich das Bild von O unter der Anwendung von $\mathbf{x}$. Umgekehrt kann man jedem Punkt P der Ebene den Vektor $\overrightarrow{OP}$ zuordnen, genannt Ortsvektor von P. Auf diese Weise erhalten wir eine umkehrbar eindeutige Zuordnung zwischen den Punkten der Ebene und den Elementen von V. Diese Zuordnung hängt natürlich von der Wahl des Punktes O ab. Ersetzt man O durch einen Punkt N, so unterscheiden sich für jeden Punkt P die Ortsvektoren bezüglich O und N um den selben Vektor $\overrightarrow{ON}$, denn

$$\overrightarrow{OP} = \overrightarrow{ON} + \overrightarrow{NP}.$$

Wählen wir eine Basis $\mathbf{v}_1$, $\mathbf{v}_2$ von V, so können wir jedem Punkt P die Koordinaten (x_1, x_2) seines Ortsvektors zuordnen. Damit haben wir ein Koordinatensystem in der Ebene definiert. Den Punkt O nennt man in diesem Zusammenhang den Koordinatenursprung. Bei der Wahl eines anderen Punktes N als Korodinatenursprung hat ein Punkt P andere Koordinaten (u_1, u_2). Sind (c_1, c_2) die Koordinaten des Punktes N bezüglich O, so hängen die verschiedenen Koordinaten von P durch die umkehrbare Substitution

$$\begin{aligned} x_1 &= c_1 + u_1 \\ x_2 &= c_2 + u_2 \end{aligned}$$

zusammen. Wechseln wir zudem noch die Basis von V, so ist auf die Koordinaten eine weitere umkehrbare (lineare) Substitution anzuwenden. Beim

Wechsel des Koordinatensystems hängen die neuen und alten Koordinaten eines Punktes also durch eine affine Substitution zusammen.

Nun sei ein Polynom $p(x_1, x_2)$ mit reellen Koeffizienten gegeben. Wir betrachten die Menge X der Punkte der Ebene, deren Koordinaten (x_1, x_2) der algebraischen Gleichung

$$p(x_1, x_2) = 0$$

genügen. Wechseln wir das Koordinatensystem, so müssen wir in p eine affine Substitution vornehmen. Dabei entsteht aus p wieder ein Polynom. Wir stellen fest, dass X in jedem Koordinatensystem durch eine algebraische Gleichung gegeben ist.

Definition 13. *Eine Menge in einer Ebene heißt algebraische Kurve, wenn sie in einem Koordinatensystem durch eine algebraische Gleichung beschrieben werden kann. Den kleinstmöglichen Grad einer solchen Gleichung nennt man die Ordnung der Kurve. Gibt es einen eindeutig bestimmten Punkt O, so dass für jeden Ortsvektor $\mathbf{x}$ eines Punktes der Kurve auch $-\mathbf{x}$ Ortsvektor eines Punktes der Kurve ist, so nennt man O den Mittelpunkt der Kurve.*

Eine Kurve zweiter Ordnung hat offenbar genau dann einen Mittelpunkt, wenn die quadratische homogene Komponente von p nicht ausgeartet ist.

Wir wollen nun die algebraischen Kurven zweiter Ordnung klassifizieren. Nach unseren obigen Ergebnissen haben ihre Gleichungen bei geeigneter Wahl des Koordinatensystems eine der folgenden Formen.

- $x_1^2 + x_2^2 = c$
  Die Kurve ist für $c > 0$ eine Ellipse, für $c = 0$ ein Punkt und für $c < 0$ die leere Menge.

- $x_1^2 - x_2^2 = c$
  Durch die umkehrbare lineare Substitution

  $$w_1 = x_1 + x_2$$
  $$w_2 = x_1 - x_2$$

  wird diese Gleichung zu

  $$w_1 w_2 = c.$$

  Die Kurve ist für $c \neq 0$ eine Hyperbel und für $c = 0$ eine Vereinigung zweier sich schneidender Geraden.

- $x_1^2 = x_2$
  Die Kurve ist eine Parabel.

- $x_1^2 = c$
  Die Kurve ist für $c > 0$ eine Vereinigung zweier paralleler Geraden, für $c = 0$ eine Gerade und für $c < 0$ die leere Menge.

Man kann noch beide Seiten dieser Gleichungen mit -1 multiplizieren, aber das führt zu keinen neuen Lösungsmengen. Man kann übrigens durch eine geeignete Substitution erreichen, dass an Stelle von c hier 1, 0 oder -1 steht. Genaugenommen müsste man noch untersuchen, welche der aufgezählten Gleichungen durch eine umkehrbare affine Substitution ineinander umgeformt werden können.

Nun wollen wir die Lösungsmengen algebraischer Gleichungen in drei Variablen geometrisch interpretieren. Dazu betrachten wir „den“ Raum und interpretieren die Menge seiner Verschiebungen als Vektorraum V. Die obige Diskussion überträgt sich auf den dreidimensionalen Fall. Wählen wir eine Basis $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ von V und einen Koordinatenursprung O im Raum, so können wir jedem Punkt seine Koordinaten (x_1, x_2, x_3) zuordnen.

Definition 14. *Eine Menge im Raum heißt algebraische Fläche, wenn sie in einem Koordinatensystem durch eine algebraische Gleichung beschrieben werden kann. Den kleinstmöglichen Grad einer solchen Gleichung nennt man die Ordnung der Fläche. Gibt es einen eindeutig bestimmten Punkt O, so dass für jeden Ortsvektor $\mathbf{x}$ eines Punktes der Fläche auch $-\mathbf{x}$ Ortsvektor eines Punktes der Fläche ist, so nennt man O den Mittelpunkt der Fläche.*

Eine Kurve zweiter Ordnung hat offenbar genau dann einen Mittelpunkt, wenn die quadratische homogene Komponente von p nicht ausgeartet ist.

Wir wollen nun die algebraischen Flächen zweiter Ordnung klassifizieren. Nach unseren obigen Ergebnissen haben ihre Gleichungen bei geeigneter Wahl des Koordinatensystems eine der folgenden Formen.

- $x_1^2 + x_2^2 + x_3^2 = c$
  Die Fläche ist für $c > 0$ ein Ellipsoid, für $c = 0$ ein Punkt und für $c < 0$ die leere Menge.

- $x_1^2 + x_2^2 - x_3^2 = c$
  Die Fläche ist für $c > 0$ ein einschaliges Hyperboloid, für $c = 0$ ein Doppelkegel und für $c < 0$ ein zweischaliges Hyperboloid (dann ist $|x_3| \geq \sqrt{-c}$ für alle seine Punkte).

- $x_1^2 + x_2^2 = x_3$
  Die Fläche ist ein elliptisches Paraboloid.

- $x_1^2 - x_2^2 = x_3$
  Durch die umkehrbare lineare Substitution

$$\begin{aligned} w_1 &= x_1 + x_2 \\ w_2 &= x_1 - x_2 \\ w_3 &= \qquad\qquad x_3 \end{aligned}$$

  wird diese Gleichung zu

$$w_1 w_2 = w_3.$$

  Die Fläche ist ein hyperbolisches Paraboloid.

- $x_1^2 + x_2^2 = c$
  Die Fläche ist für $c > 0$ ein elliptischer Zylinder, für $c = 0$ eine Gerade und für $c < 0$ die leere Menge.

- $x_1^2 - x_2^2 = c$
  Die Fläche ist für $c \neq 0$ ein hyperbolischer Zylinder und für $c = 0$ eine Vereinigung zweier sich schneidender Ebenen.

- $x_1^2 = x_2$
  Die Fläche ist ein parabolischer Zylinder.

- $x_1^2 = c$
  Die Fläche ist für $c > 0$ eine Vereinigung zweier paralleler Ebenen, für $c = 0$ eine Ebene und für $c < 0$ die leere Menge.

Die nach der Klassifikation von Kurven gemachten Bemerkungen gelten hier ebenfalls. Bilder der Flächen zweiter Ordnung findet man z. B. auf der Seite von M. Stroppel. Auch für größeres n interpretiert man die Lösungsmengen von algebraischen Gleichungen in n Variablen geometrisch als sogenannte Hyperflächen im n-dimensionalen reellen affinen Raum, wenngleich die Anschauung dann versagt. Eine (Hyper-)Fläche zweiter Ordnung nennt man in neuerer Zeit auch Quadrik. Was aber ist ein affiner Raum?

Definition 15. *Es sei A eine Menge, deren Elemente wir Punkte nennen, bei der für jedes Quadrupel von Punkten (P, Q, R, S) festgelegt ist, ob sie ein Parallelogramm bilden. Wenn ja, schreiben wir $P \overset{Q}{\underset{R}{}} S$. Die Menge A heißt affiner Raum, wenn folgendes gilt.*

(a) Für beliebige Punkte P, Q und R gibt es genau einen Punkt S, so dass $P \overset{Q}{\underset{R}{}} S$.

(b) Es gilt immer $P \overset{P}{\underset{Q}{}} Q$.

(c) Es gilt genau dann $P \overset{Q}{\underset{R}{}} S$, wenn $P \overset{R}{\underset{Q}{}} S$.

(d) Es gilt genau dann $P \overset{Q}{R} S$, *wenn* $R \overset{S}{P} Q$.

(e) Wenn $P \overset{Q}{R} S$ *und* $R \overset{S}{T} U$, *dann* $P \overset{Q}{T} U$.

Eine Teilmenge B eines affinen Raumes A heißt affiner Unterraum, wenn die Menge B mit der Einschränkungen der Parallelogramm-Relation ein affiner Raum ist.

Für gegebene Punkte P und Q definieren wir eine Abbildung von A in sich selbst, genannt Verschiebung $\overrightarrow{PQ}$, wie folgt. Ist R ein beliebiger Punkt, so ist das Bild von R derjenige Punkt S, für den $P \overset{Q}{R} S$. Nach (b) und (c) ist das Bild von P unter $\overrightarrow{PQ}$ gleich Q. Nach (d) und (e) ist eine Verschiebung, die P auf Q abbildet, gleich $\overrightarrow{PQ}$. Die Nacheinanderausführung von Verschiebungen (erst $\mathbf{x}$ und dann $\mathbf{y}$) wird mit $\mathbf{x}+\mathbf{y}$ bezeichnet. Dann gilt

$$\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}.$$

Satz 5. *Es sei A ein affiner Raum und V die Menge seiner Verschiebungen.*

(i) Die identische Abbildung (die wir mit $\mathbf{0}$ bezeichnen) gehört zu V.

(ii) Für alle $\mathbf{x}$, $\mathbf{y} \in V$ gilt $\mathbf{x}+\mathbf{y} = \mathbf{y}+\mathbf{x}$

(iii) Für alle $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z} \in V$ gilt $(\mathbf{x}+\mathbf{y})+\mathbf{z} = \mathbf{x}+(\mathbf{y}+\mathbf{z})$.

(iv) Für jedes $\mathbf{x} \in V$ gibt es ein $\mathbf{u} \in V$ (den entgegengesetzten Vektor), so dass $\mathbf{x}+\mathbf{u} = \mathbf{0}$.

(v) Ist B ein affiner Unterraum und W die Menge der Verschiebungen $\overrightarrow{PQ}$ für Punkte P, $Q \in B$, so ist W abgeschlossen unter $+$ und enthält $\mathbf{0}$ sowie zu jedem seiner Vektoren den entgegengesetzten Vektor.

Beweis. Es sei P ein beliebiger Punkt.

(i) Nach (a) ist $\overrightarrow{PP}$ die identische Abbildung.

(ii) Es sei Q das Bild von P unter $\mathbf{x}$ und R das Bild von P unter $\mathbf{y}$, und es sei S der Punkt mit der Eigenschaft $P \overset{Q}{R} S$. Dann ist S das Bild von R unter $\mathbf{x}$, also das Bild von P unter $\mathbf{y}+\mathbf{x}$. Nach (c) gilt $P \overset{R}{Q} S$, also ist S das Bild von Q unter $\mathbf{y}$ und somit das Bild von P unter $\mathbf{x}+\mathbf{y}$.

(iii) folgt aus allgemeinen Eigenschaften von Abbildungen.

(iv) Es sei Q das Bild von P unter $\mathbf{x}$. Dann ist $\mathbf{x} = \overrightarrow{PQ}$, und

$$\overrightarrow{PQ} + \overrightarrow{QP} = \overrightarrow{PP} = \mathbf{0}.$$

(v) folgt direkt aus den Definitionen und dem Bewiesenen. □

Beispiel. Es sei A eine Kreislinie. Für Punkte P, Q, R und S von A sagen wir, dass $P \overset{Q}{R} S$, wenn die Drehung um den Mittelpunkt von A, welche P in Q überführt, auch R in S überführt. Dann ist A ein affiner Raum im Sinne der obigen Definition, wobei die Drehungen als Verschiebungen bezeichnet werden.

Wir müssen also den Begriff des affinen Raumes noch einschränken, um das zu erhalten, was wir uns intuitiv darunter vorstellen.

Definition 16. *Ein reeller affiner Raum ist ein affiner Raum A, bei dem für jede Verschiebung $\mathbf{x}$ und jede reelle Zahl a eine Verschiebung $a \cdot \mathbf{x}$ definiert ist, so dass die Menge der Verschiebungen V zu einem reellen Vektorraum wird. Die Dimension von A ist die Dimension von V.*

Ein reeller affiner Unterraum von A ist ein affiner Unterraum B, dessen Menge von Verschiebungen abgeschlossen unter der Skalarmultiplikation ist.

Bisher haben wir Geraden und Ebenen im Raum intuitiv betrachtet. Streng genommen handelt es sich um eindimensionale und zweidimensionale reelle affine Unterräume in einem dreidimensionalen reellen affinen Raum.

Es sei beispielsweise B eine Ebene im Raum A und W die Menge der Verschiebungen von A, die alle Punkte von B in Punkte von B überführen. Wir können einen Koordinatenursprung N in B und eine Basis $\mathbf{w}_1$, $\mathbf{w}_2$ von W wählen. Dann besteht die Ebene B aus allen Punkten P, für die

$$\overrightarrow{NP} = t_1\mathbf{w}_1 + t_2\mathbf{w}_2$$

mit geeigneten reellen Zahlen t_1, t_2 ist. Andererseits kann man einen Koordinatenursprung $O \in A$ und eine Basis $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$ des Vektorraums V der Verschiebungen von A wählen, so dass

$$\overrightarrow{OP} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3$$

mit geeigneten reellen Zahlen x_1, x_2, x_3 ist. Da hier zwei Koordinatensysteme im Spiel sind, bezeichnet man der Übersicht halber die Zahlen t_1, t_2 nicht als Koordinaten, sondern als Parameter des Punktes P auf der Ebene B. Drücken wir alle Vektoren durch die Basisvektoren von V aus, also

$$\begin{aligned}
\mathbf{w}_1 &= b_{11}\mathbf{v}_1 + b_{21}\mathbf{v}_2 + b_{31}\mathbf{v}_3 \\
\mathbf{w}_2 &= b_{12}\mathbf{v}_1 + b_{22}\mathbf{v}_2 + b_{32}\mathbf{v}_3 \\
\overrightarrow{ON} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3
\end{aligned}$$

mit reellen Zahlen b_{ij} und c_i, so ergibt sich durch Einsetzen in die Gleichung $\overrightarrow{OP} = \overrightarrow{ON} + \overrightarrow{NP}$ ähnlich wie auf S. 18 die so genannte Parametrisierung

$$\begin{aligned}
x_1 &= b_{11}t_1 + b_{12}t_2 + c_1 \\
x_2 &= b_{21}t_1 + b_{22}t_2 + c_2 \\
x_3 &= b_{31}t_1 + b_{32}t_2 + c_3
\end{aligned}$$

Um die Schnittkurve der Ebene B mit einer algebraischen Fläche zu bestimmen, müssen wir diese Parametrisierung in die Gleichung einsetzen, welche die Fläche im Koordinatensystem x_1, x_2, x_3 beschreibt, und die entstehende Gleichung in den Variablen t_1, t_2 untersuchen.

8 Skalarprodukte und Normen

Definition 17. *Ein Skalarprodukt auf einem endlichdimensionalen reellen Vektorraum V ist eine positiv definite Bilinearform b auf V. Die Norm eines Vektors $\mathbf{x} \in V$ bezüglich b ist $\|\mathbf{x}\| = \sqrt{b(x,x)}$.*

Offensichtlich gilt für $t \in \mathbf{R}$

$$\|t \cdot \mathbf{x}\| = |t| \cdot \|\mathbf{x}\|.$$

Satz 6 (Cauchy-Schwarz-Ungleichung)**.** *Für beliebige Vektoren $\mathbf{x}$, $\mathbf{y}$ in einem Vektorraum V mit Skalarprodukt gilt*

$$|b(\mathbf{x},\mathbf{y})| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

Dieser Satz wurde zuerst von Augustin Louis Cauchy (1789–1857) bewiesen. Die Verallgemeinerung durch Hermann Amandus Schwarz (1843–1921) wurde eigentlich schon vorher von Wiktor Jakowlewitsch Bunjakowski (1804–1889) gefunden.

Beweis. Ist einer der Vektoren gleich $\mathbf{0}$, so wird die Behauptung zu $0 = 0$. Also sei $\mathbf{x} \neq \mathbf{0}$. Da b positiv definit ist, gilt für die Spezialisierung q von b

$$q(q(\mathbf{x})\mathbf{y} - b(\mathbf{x},\mathbf{y})\mathbf{x}) \geq 0.$$

Wegen der Bilinearität von b bedeutet dies

$$q(\mathbf{x})^2 q(\mathbf{y}) - 2q(\mathbf{x})b(\mathbf{x},\mathbf{y})^2 + q(\mathbf{x})b(\mathbf{x},\mathbf{y})^2 \geq 0,$$

also nach Division durch $q(\mathbf{x})$

$$q(\mathbf{x})q(\mathbf{y}) \geq b(\mathbf{x},\mathbf{y})^2.$$

Nun folgt die Behauptung aus der Monotonie der Wurzelfunktion für nichtnegative Argumente. □

Folgerung 2. *In der Situation des Satzes gilt*

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|.$$

Es ist nämlich

$$q(\mathbf{x}+\mathbf{y}) = q(\mathbf{x}) + 2b(\mathbf{x},\mathbf{y}) + q(\mathbf{y}) \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2.$$

Nun sei A ein affiner Raum und V der Vektorraum der Verschiebungen von A. Wählt man auf V eine Norm, so kann den Abstand zweier Punkte P und Q von A durch

$$d(P,Q) = \|\overrightarrow{PQ}\|$$

definieren. Man bezeichnet A zusammen mit dieser Abstandsfunktion d als Euklidischen Raum. Die Folgerung aus Satz 6 kann man nun auch so formulieren, dass für beliebige Punkte P, Q und R von A gilt

$$d(P,Q) + d(Q,R) \leq d(P,R).$$

Dies (und auch die Aussage der Folgerung) bezeichnet man als Dreiecksungleichung.

Mit Hilfe der Abstandsfunktion können wir folgende Begriffe einführen, die wir später benötigen werden.

Definition 18. *Eine Funktion f auf einer Teilmenge X von A heißt dehnungsbeschränkt, wenn es eine positive Zahl C gibt, so dass für alle $P, Q \in X$ gilt*

$$|f(P) - f(Q)| \leq C \cdot d(P,Q).$$

Offenbar ist jede polynomiale Funktion vom Grad höchstens 1 dehnungsbeschränkt. Bei Polynomen höheren Grades ist das nur der Fall, wenn man sie auf eine genügend kleine Menge einschränkt.

Definition 19. *Eine Teilmenge U von A heißt Umgebung des Punktes P, wenn es eine positive Zahl e gibt, so dass alle Punkte Q mit der Eigenschaft $d(P,Q) < e$ zu U gehören.*

Auf den ersten Blick scheinen die Begriffe der Umgebung und der Dehnungsbeschränktheit von der Wahl der Norm abzuhängen. Dies ist aber angesichts von Aufgabe 25 nicht der Fall.

9 Differentiale

Es sei f eine Funktion auf einem reellen affinen Raum A mit Werten im Körper $\mathbf{R}$ der reellen Zahlen. Grob gesagt heißt die Funktion f differenzierbar in einem Punkt P von A, wenn sie sich in einer Umgebung dieses Punktes gut durch ein Polynom vom Grad 1 annähern lässt. Die genaue Definition benötigt den Begriff des Grenzwertes und soll hier nicht gegeben werden. Statt dessen begnügen wir uns mit einem einfacheren Differenzierbarkeitsbegriff. Dazu wählen wir eine Norm auf dem Vektorraum V der Verschiebungen und bezeichnen die zugehörige Abstandsfunktion auf A mit d.

Definition 20. *Es sei X eine Teilmenge eines affinen Raumes A und P ein Punkt von X. Eine Funktion $f : X \to \mathbf{R}$ heißt stark differenzierbar an der Stelle P, wenn es eine polynomiale Funktion p vom Grad höchstens 1 auf V, eine Umgebung U von P in X und eine positive Zahl C gibt, so dass für alle $Q \in U$ gilt*

$$|f(Q) - p(Q)| \leq C \cdot d(P, Q)^2.$$

Bezeichnen wir die Differenzfunktion $f - p$ mit r, so können wir die Bedingung auch in der Form

$$f = p + r, \qquad |r(Q)| \leq C \cdot d(P, Q)^2$$

schreiben. Da ein Punkt Q eindeutig durch den Vektor $\overrightarrow{PQ}$ bestimmt ist und umgekehrt, können wir p auch als Funktion von $\overrightarrow{PQ}$ auffassen und dann in homogene Komponenten zerlegen. Es ist also

$$p(Q) = c + l(\overrightarrow{PQ}),$$

wobei c eine Konstante und l eine Linearform auf V ist.

Lemma 3. *Das Polynom in der obigen Definition ist eindeutig bestimmt.*

Das gilt dann natürlich auch für die Linearform l. Man nennt sie das Differential von f an der Stelle P.

Beweis. Setzen wir $Q = P$, so erhalten wir

$$|r(P)| \leq 0, \qquad f(P) = p(P), \qquad p(P) = c,$$

also ist $c = f(P)$ eindeutig bestimmt. Gibt es ein weiteres Polynom vom Grad 1

$$q(Q) = c + m(\overrightarrow{PQ})$$

und eine Konstante $D > 0$, so dass ebenfalls

$$f = q + s, \qquad |s(Q)| \leq D \cdot d(P, Q)^2,$$

so folgt nach der Dreiecksungleichung

$$|q(Q) - p(Q)| = |r(Q) - s(Q)| \leq |r(Q)| + |s(Q)| \leq (C + D)d(P, Q)^2.$$

Setzen wir die Formeln für p und q ein, so kürzt sich c weg, und wir können alles durch den Vektor $\overrightarrow{PQ}$ ausdrücken, den wir mit $\mathbf{t}$ bezeichnen. Wählen

wir die positive Zahl e wie in der Definition der Umgebung, so ergibt sich für alle $\mathbf{t} \in V$ mit der Eigenschaft $\|\mathbf{t}\| < e$, dass

$$|l(\mathbf{t}) - m(\mathbf{t})| \leq (C + D)\|\mathbf{t}\|^2.$$

Hier können wir $\mathbf{t}$ durch $s\mathbf{t}$ ersetzen, wobei $0 \leq s \leq 1$, und wegen der Homogenität von l und m folgt

$$s|l(\mathbf{t}) - m(\mathbf{t})| \leq s^2(C + D)\|\mathbf{t}\|^2,$$

und für $0 < s \leq 1$

$$|l(\mathbf{t}) - m(\mathbf{t})| \leq s(C + D)\|\mathbf{t}\|^2.$$

Angenommen, die linke Seite ist für einen Vektor $\mathbf{t}$ nicht Null. Dann muss $\mathbf{t} \neq \mathbf{0}$ sein, also $\|\mathbf{t}\| > 0$ wegen der Definition der Norm, und wir können beide Seiten durch $(C + D)\|\mathbf{t}\|^2$ dividieren. Setzen wir s gleich der kleineren der beiden Zahlen

$$\frac{|l(\mathbf{t}) - m(\mathbf{t})|}{2(C + D)\|\mathbf{t}\|^2} \quad \text{und} \quad 1,$$

so erhalten wir einen Widerspruch. Für $\|\mathbf{t}\| \leq e$ folgt also $l(\mathbf{t}) = m(\mathbf{t})$, und wegen der Homogenität gilt dies für alle $\mathbf{t} \in V$. □

Die Begriffe der starken Differenzierbarkeit und des Differentials hängen angesichts von Aufgabe 25 nicht von der Wahl der Norm ab. Die starke Differenzierbarkeit ist eine lokale Eigenschaft einer Funktion: Stimmen zwei Funktionen in einer Umgebung von P überein, so haben sie auch das selbe Differential an der Stelle P. Eine wichtige Anwendung ist durch folgenden Satz gegeben.

Satz 7. *Ist die Funktion f an der Stelle P stark differenzierbar und hat dort ein lokales Extremum, so ist P ein stationärer Punkt von f, das heißt, das Differential von f an dieser Stelle ist gleich Null.*

Beweis. Angenommen, f hat an der Stelle P ein lokales Minimum, das heißt, es gibt eine positive Zahl e, so dass für $d(P, Q) \leq e$ gilt $f(P) \leq f(Q)$. Schreiben wir $f(Q) = f(P) + l(\overrightarrow{PQ}) + r(Q)$, wobei l das Differential ist, so folgt

$$-l(\overrightarrow{PQ}) \leq r(Q).$$

Wenn wir e klein genug wählen, so gibt es eine Konstante $C > 0$, so dass für $\mathbf{t} \in V$ mit der Eigenschaft $\|\mathbf{t}\| \leq e$ gilt

$$-l(\mathbf{t}) \leq C\|\mathbf{t}\|^2.$$

Ersetzen wir $\mathbf{t}$ durch $-\mathbf{t}$, so erhalten wir

$$l(\mathbf{t}) \leq C\|\mathbf{t}\|^2,$$

zusammengefasst also

$$|l(\mathbf{t})| \leq C\|\mathbf{t}\|^2.$$

Für $0 < s \leq 1$ folgt, indem wir $\mathbf{t}$ durch $s\mathbf{t}$ ersetzen und durch s dividieren,

$$|l(\mathbf{t})| \leq sC\|\mathbf{t}\|^2.$$

Wäre $\mathbf{t} \neq \mathbf{0}$, aber $l(\mathbf{t}) \neq 0$, so könnten wir s gleich der kleineren der beiden Zahlen

$$\frac{|l(\mathbf{t})|}{2\|\mathbf{t}\|^2} \quad \text{und} \quad 1$$

setzen und erhielten einen Widerspruch. Also ist $l(\mathbf{t}) = 0$, falls $\|\mathbf{t}\| \leq e$, und wegen der Homogenität gilt das für beliebige $\mathbf{t}$.

Der Beweis für ein lokales Maximum ist ähnlich, kann aber auch durch Anwendung des Bewiesenen auf die Funktion $-f$ ersetzt werden. □

Es ist offensichtlich, dass das Differential einer polynomialen Funktion vom Grad höchstens 1 gleich der linearen homogenen Komponente des Polynoms ist. Um die Differentiale weiterer Funktionen praktisch zu bestimmen, benutzt man Differentiationsregeln wie die folgenden.

Satz 8. *Die Funktionen f und g auf einer Teilmenge X eines affinen Raumes A seien an der Stelle P stark differenzierbar, und ihre Differentiale seien l und m. Dann sind auch die Funktionen $f + g$ und $f \cdot g$ stark differenzierbar, und ihre Differentiale sind $l + m$ bzw. $g(P) \cdot l + f(P) \cdot m$. Ist zudem $g(P) \neq 0$, so ist auch die Funktion $\frac{f}{g}$ an der Stelle P stark differenzierbar.*

Man kann das Differential von $\frac{f}{g}$, nennen wir es n, wie folgt bestimmen. Wegen $g \cdot \frac{f}{g} = f$ ist nach der Produktregel aus dem Satz

$$\frac{f(P)}{g(P)} \cdot m + g(P) \cdot n = l,$$

also

$$n = \frac{1}{g(P)} \cdot l - \frac{f(P)}{g(P)^2} \cdot m.$$

Beweis von Satz 8. Wir beweisen zunächst die Produktregel. Laut Definition gilt

$$f = p + r, \qquad g = q + s,$$

wobei

$$p(Q) = f(P) + l(\overrightarrow{PQ}), \qquad q(Q) = g(P) + m(\overrightarrow{PQ})$$

und es Konstanten B, C und e gibt, so dass für $d(P,Q) \leq e$ gilt

$$|r(Q)| \leq B \cdot d(P,Q)^2, \qquad |s(Q)| \leq C \cdot d(P,Q)^2.$$

Wegen der Dehnungsbeschränktheit von Linearformen (Aufgabe 26) gibt es Konstanten D und E, so dass

$$|l(\overrightarrow{PQ})| \leq D \cdot d(P,Q), \qquad |m(\overrightarrow{PQ})| \leq E \cdot d(P,Q),$$

also nach der Dreiecksungleichung

$$|p(Q)| \leq |f(P)| + Be, \qquad |q(Q)| \leq |g(P)| + Ce.$$

Wir bezeichnen die rechten Seiten mit F und G.

Wir untersuchen nun die Produktfunktion

$$f \cdot g = p \cdot q + p \cdot s + q \cdot r + r \cdot s.$$

Für $d(P,Q) \leq e$ gilt

$$|p(Q) \cdot s(Q) + q(Q) \cdot r(Q) + r(Q) \cdot s(Q)| \leq (FE + GD + e^2) d(P,Q)^2.$$

Außerdem ist

$$p(Q)q(Q) = f(P)g(P) + g(P)l(\overrightarrow{PQ}) + f(P)m(\overrightarrow{PQ}) + l(\overrightarrow{PQ})m(\overrightarrow{PQ}),$$

wobei der letzte Term wegen

$$|l(\overrightarrow{PQ})m(\overrightarrow{PQ})| \leq DE \cdot d(P,Q)^2$$

dem Restglied zugeschlagen werden kann. Mit der Definition des Differentials folgt die Behauptung.

Der Beweis der Summenregel ist ähnlich, aber viel einfacher.

Die starke Differenzierbarkeit von Quotientenfunktionen brauchen wir nur für den Zähler 1 zu beweisen, weil der allgemeine Fall dann mit der Produktregel folgt. Wir müssen also zeigen, dass $\frac{1}{g(Q)}$ in einer Umgebung von P gut durch

$$\frac{1}{g(P)} - \frac{1}{g(P)^2} m(\overrightarrow{PQ})$$

approximiert wird, genauer gesagt, dass der Absolutbetrag der Differenz

$$\frac{1}{g(P)} - \frac{1}{g(Q)} - \frac{1}{g(P)^2} m(\overrightarrow{PQ}) \tag{7}$$

für $d(P,Q) \leq e$ durch ein Vielfaches von $d(P,Q)^2$ beschränkt ist. Nun ist aber

$$\frac{1}{g(P)} - \frac{1}{g(Q)} = \frac{g(Q) - g(P)}{g(P)g(Q)} = \frac{m(\overrightarrow{PQ}) + s(Q)}{g(P)g(Q)}.$$

Nach der Dreiecksungleichung ist

$$|g(P)| \leq |g(Q)| + |m(\overrightarrow{PQ})| \leq |g(Q)| + Ee.$$

Da die Zahl $a = |g(P)|$ positiv ist, können wir e so klein wählen, dass auch die Zahl $b = a - Ee$ positiv ist, und für $d(P,Q) \leq e$ folgt

$$|g(Q)| \geq b, \qquad \left| \frac{1}{g(P)} - \frac{1}{g(Q)} \right| \leq \frac{E + Ce}{ab} d(P,Q).$$

Andererseits können wir den Ausdruck (7) in der Form

$$\frac{m(\overrightarrow{PQ}) + s(Q)}{g(P)g(Q)} - \frac{1}{g(P)^2} m(\overrightarrow{PQ}) = \frac{m(\overrightarrow{PQ})}{g(P)} \left(\frac{1}{g(Q)} - \frac{1}{g(P)} \right) + \frac{s(Q)}{g(P)g(Q)}$$

schreiben, und sein Absolutbetrag ist beschränkt durch

$$\left(\frac{E(E + Ce)}{a^2 b} + \frac{C}{ab} \right) d(P,Q)^2. \qquad \square$$

Für praktische Rechnungen wählt man meist ein Koordinatensystem auf dem affinen Raum A, also einen Koordinatenursprung O und eine Basis $\mathbf{v}_1$, $\ldots$, $\mathbf{v}_n$ von V. Da uns die Wahl der Norm auf V freisteht, wählen wir der Einfachheit halber

$$\|t_1 \mathbf{v}_1 + \ldots + t_n \mathbf{v}_n\| = \sqrt{t_1^2 + \ldots + t_n^2}.$$

Ein Punkt P ist durch seine Koordinaten $x_1, \ldots, x_n$ gegeben, und statt $f(P)$ schreiben wir $f(x_1, \ldots, x_n)$. Das Differential l von f an der Stelle P ist dann durch die Zahlen

$$a_i = l(\mathbf{v}_i)$$

bestimmt. Wie kann man sie ermitteln? Bezeichnen wir die Koordinaten von $\overrightarrow{PQ}$ mit t_i, so ist laut Definition des Differentials

$$f(x_1 + t_1, \ldots, x_n + t_n) = f(x_1, \ldots, x_n) + a_1 t_1 + \ldots + a_n t_n + r(t_1, \ldots, t_n),$$

wobei

$$|r(t_1, \ldots, t_n)| \leq C(t_1^2 + \ldots + t_n^2)$$

für genügend kleine Werte von $t_1^2 + \ldots + t_n^2$.

Betrachten wir zunächst den Fall $n = 1$. Dann brauchen wir keine Indizes, und unsere Bedingung wird zu

$$f(x+t) = f(x) + at + r(t), \qquad |r(t)| \leq Ct^2$$

für kleine $|t|$. Die Zahl a nennt man dann die Ableitung von f an der Stelle x und bezeichnet sie nach Joseph-Louis Lagrange (1736–1813) mit $f'(x)$. Aus Satz 8 folgen die bekannten Ableitungsregeln

$$(f+g)' = f' + g', \qquad (f \cdot g)' = f' \cdot g + f \cdot g', \qquad \left(\frac{f}{g}\right)' = \frac{f' \cdot g - f \cdot g'}{g^2},$$

die gelten, sobald die rechten Seiten definiert sind. Des weiteren gilt die Kettenregel: Ist f an der Stelle x stark differenzierbar und ist g an der Stelle $y = f(x)$ stark differenzierbar, so ist die durch $g \circ f(x) = g(f(x))$ definierte Verkettung $f \circ g$ an der Stelle x stark differenzierbar, und es gilt

$$(g \circ f)'(x) = g'(y) \cdot f'(x).$$

Betrachtet man die Ableitungen wieder als Funktionen, so kann man die Kettenregel auch in der Form

$$(g \circ f)' = (g' \circ f) \cdot f'$$

schreiben.

Die Bezeichnung „Differential“ geht auf Leibniz (1646–1716) zurück, der darunter allerdings eine unendlich kleine Größe verstand, die es in der Standardmathematik nicht gibt. Später hat man das Differential df der Funktion f an der Stelle P als die o. g. Linearform l umgedeutet. Man kann die Variable x als Funktion auf $\mathbf{R}$ interpretieren, die somit ebenfalls ein Differential besitzt, und dann gilt

$$df = f'(x)\, dx,$$

was Leibniz' Schreibweise $\frac{df}{dx}$ (gelesen „d f nach d x“) und die Bezeichnung „Differentialquotient“ für die Ableitung halbwegs rehabilitiert. Traditionell ist es allerdings unüblich, die Abhängigkeit des Differentials df vom Punkt x und der Variablen t auszudrücken.

Nun betrachten wir einen affinen Raum A von beliebiger Dimension n. Ist von den Koordinaten t_i nur eine von Null verschieden, so wird die Bedingung an das Differential zu

$$f(x_1, \ldots, x_i + t_i, \ldots, x_n) = f(x_1, \ldots, x_n) + a_i t_i + r(0, \ldots, t_i, \ldots, 0),$$

wobei

$$|r(0,\ldots,t_i,\ldots,0)| \leq C \cdot t_i^2$$

für kleine $|t_i|$. Wir sehen, dass a_i nichts anderes ist als die Ableitung von f als Funktion der $i$ten Variablen, wobei alle anderen Variablen konstant gehalten werden. Dies nennt man die partielle Ableitung von f nach der $i$ten Variablen. Gegen die klassische Schreibweise $\frac{\partial f}{\partial x_i}$ (gelesen „d f nach d x i“) von Gustav Jacobi (1804–1851) konnte sich keine einheitliche Schreibweise durchsetzen, welche die Abhängigkeit von $(x_1,\ldots,x_n)$ zum Ausdruck bringt.

Auch im mehrdimensionalen Fall ist die Bezeichnung df für das Differential der Funktion f üblich. Hier kann man die Variable x_i als Funktion interpretieren, die einem Punkt seine $i$te Koordinate zuordnet. Dann ist das Differential dx_i eine Linearform, die auf $\mathbf{v}_i$ den Wert 1 und auf allen anderen Basisvektoren den Wert 0 annimmt. Somit gilt

$$df = \frac{\partial f}{\partial x_1} dx_1 + \ldots + \frac{\partial f}{\partial x_n} dx_n.$$

Man kann die Ableitungsregeln aus Satz 8 in der klassischen Schreibweise von Differentialen ausdrücken:

$$d(f+g) = df + dg, \qquad d(f \cdot g) = df \cdot g + f \cdot dg, \qquad d\left(\frac{f}{g}\right) = \frac{df \cdot g - f \cdot dg}{g^2}.$$

Beispiel. Wir betrachten die Funktion

$$f(x_1,x_2) = 3\sqrt{x_1^2 + x_2^2 + 1} - x_1 - 2x_2.$$

Dann sind die partiellen Ableitungen gegeben durch

$$\frac{\partial f}{\partial x_1} = \frac{3x_1}{\sqrt{x_1^2 + x_2^2 + 1}} - 1,$$
$$\frac{\partial f}{\partial x_2} = \frac{3x_2}{\sqrt{x_1^2 + x_2^2 + 1}} - 2.$$

Hier haben wir neben den bereits genannten Ableitungsregeln auch die Kettenregel benutzt. An einem stationären Punkt müssen die partiellen Ableitungen verschwinden, d. h.

$$\frac{3x_1}{\sqrt{x_1^2 + x_2^2 + 1}} = 1$$
$$\frac{3x_2}{\sqrt{x_1^2 + x_2^2 + 1}} = 2$$

Es folgt $2x_1 = x_2$. Damit können wir x_2 aus der ersten Gleichung eliminieren. Multiplizieren wir beide Seiten mit dem positiven Nenner, so folgt

$$3x_1 = \sqrt{5x_1^2 + 1}.$$

Quadrieren ergibt $9x_1^2 = 5x_1^2 + 1$, also $x_1^2 = \frac{1}{4}$. Als Quadratwurzel muss $3x_1$ positiv sein. Die Funktion f hat somit nur den stationären Punkt $\left(\frac{1}{2}, 1\right)$.

Strenggenommen muss man zunächst prüfen, ob eine gegebene Funktion tatsächlich stark differenzierbar ist. Dies geschieht praktisch meist unter Benutzung des folgenden Kriteriums.

Satz 9. *Es sei f eine Funktion in der Umgebung eines Punktes P in einem affinen Raum, auf dem ein Koordinatensystem gewählt ist. Wenn f in einer Umgebung von P bezüglich jeder Koordinate stark partiell differenzierbar ist und die partiellen Ableitungen dehnungsbeschränkt sind, dann ist f an der Stelle P stark differenzierbar.*

Damit ist die Frage auf die starke Differenzierbarkeit von Funktionen einer Variablen zurückgeführt. Die elementaren Funktionen, die man aus Konstanten, der identischen Funktion, den Winkelfunktionen, der Exponentialfunktion und der Logarithmusfunktion durch Rechenoperationen und Verkettung bilden kann, sind auf ihren natürlichen Definitionsbereichen sämtlich stark differenzierbar.

10 Zweite Differentiale

Bisher haben wir das Differential $l(\mathbf{t})$ einer Funktion f nur an einer festen Stelle P betrachtet. Wenn wir sagen, dass f auf einer Menge X stark differenzierbar ist, so meinen wir, dass dies an jeder Stelle Q von X der Fall ist. In diesem Fall hängt das Differential natürlich von Q ab, und wir schreiben $l(\mathbf{t}, Q)$.

Definition 21. *Es sei X eine Teilmenge eines affinen Raumes A und P ein Punkt von X. Eine Funktion $f : X \to \mathbf{R}$ heiÃßt zweimal stark differenzierbar an der Stelle P, wenn sie an jeder Stelle Q einer Umgebung U von P stark differenzierbar ist und ihr Differential für jeden festen Vektor als Funktion von Q wiederum stark differenzierbar an der Stelle P ist.*

Für festes $\mathbf{t}$ ist das Differential der Funktion $l(\mathbf{t}, Q)$ an der Stelle P eine Linearform, deren Argument wir mit einer neuen Variablen $\mathbf{u}$ bezeichnen. Außerdem hängt es immer noch von $\mathbf{t}$ ab, und wir bezeichnen es mit $h(\mathbf{t}, \mathbf{u})$. Man nennt es das zweite Differential von f an der Stelle P.

Ausführlich bedeutet die starke Differenzierbarkeit von f an einer Stelle Q, dass es eine Zahl $C_1(Q)$ gibt, so dass für R in einer Umgebung von Q gilt

$$f(R) = f(Q) + l(\overrightarrow{QR}, Q) + r_1(Q, R), \qquad |r_1(Q, R)| \leq C_1(Q) d(Q, R)^2.$$

Für festes $\mathbf{t}$ bedeutet die starke Differenzierbarkeit von l an der Stelle P, dass es eine Zahl $C_2(\mathbf{t})$ gibt, so dass für Q in einer Umgebung von P gilt

$$l(\mathbf{t}, Q) = l(\mathbf{t}, P) + h(\mathbf{t}, \overrightarrow{PQ}) + r_2(\mathbf{t}, Q), \qquad |r_2(\mathbf{t}, Q)| \leq C_2(\mathbf{t}) d(P, Q)^2.$$

Wegen

$$l(\mathbf{s} + \mathbf{t}, Q) = l(\mathbf{s}, Q) + l(\mathbf{t}, Q), \qquad l(a \cdot \mathbf{t}, Q) = a \cdot l(\mathbf{t}, Q)$$

für alle $\mathbf{s}$, $\mathbf{t} \in V$, $a \in \mathbf{R}$ und $Q \in X$ folgt mit Satz 8, dass

$$h(\mathbf{s} + \mathbf{t}, \mathbf{u}) = h(\mathbf{s}, \mathbf{u}) + h(\mathbf{t}, \mathbf{u}), \qquad h(a \cdot \mathbf{t}, \mathbf{u}) = a \cdot h(\mathbf{t}, \mathbf{u}).$$

Somit ist das zweite Differential eine Bilinearform, die man nach Otto Hesse (1811–1874) auch Hesse-Form[6] nennt.

Satz 10 (Schwarz)**.** *Ist f in einer Umgebung von P zweimal stark differenzierbar und ist das zweite Differential $h(\mathbf{t}, \mathbf{u})$ für feste $\mathbf{t}$ und $\mathbf{u}$ in einer Umgebung von P dehnungsbeschränkt, so ist es im Punkt P eine symmetrische Bilinearform.*

Wir weichen hier etwas von der ursprünglichen Formulierung durch Hermann Amandus Schwarz ab.

Satz 11. *Die Funktion f sei auf einer Teilmenge X des affinen Raumes A zweimal stark differenzierbar, und ihr zweites Differential sei dehnungsbeschränkt. Dann gibt es für jeden Punkt P von X eine polynomiale Funktion p vom Grad höchstens 2 und eine Konstante C, so dass*

$$|f(Q) - p(Q)| \leq C \cdot d(P, Q)^3.$$

Ist l das (erste) Differential und h das zweite Differential von f an der Stelle P, und ist q die Spezialisierung von h, so gilt

$$p(Q) = f(P) + l(\overrightarrow{PQ}) + \frac{1}{2} q(\overrightarrow{PQ}).$$

Man nennt p das Taylor-Polynom zweiten Grades von f an der Stelle P nach Brook Taylor (1685–1731), der eine solche Formel mit Polynomen beliebigen Grades entdeckt hat. Die Beweise der Sätze 9, 10 und 11 benutzen den Mittelwertsatz der Differentialrechnung und werden hier nicht gegeben.

[6] Hesse hatte behauptet, dass eine Form, deren zweites Differential an jeder Stelle ausgeartet ist, sich durch umkehrbare lineare Substitutionen auf eine Form zurückführen lässt, in der eine der Variablen fehlt. Obwohl sich dies als falsch herausstellte, blieb es bei der Bezeichnung.

Satz 12. *In der Situation von Satz 11 sei P ein stationärer Punkt von f. Dann gilt:*

(i) Ist h positiv definit, so hat f an der Stelle P ein lokales Minimum.

(ii) Ist h negativ definit, so hat f an der Stelle P ein lokales Maximum.

(iii) Ist h indefinit, so hat f an der Stelle P kein lokales Extremum.

Beweis. Für $d(P,Q) < e$ haben wir

$$f(Q) = f(P) + \frac{1}{2}q(\overrightarrow{PQ}) + r(\overrightarrow{PQ}),$$

wobei für $\|\mathbf{t}\| < e$ gilt

$$|r(\mathbf{t})| \leq C\|\mathbf{t}\|^3.$$

Ist $0 < s < 1$ und R das Bild von P unter der Verschiebung $s\mathbf{t}$, so ist $d(P,R) < e$, und ist außerdem

$$s < \frac{|q(\mathbf{t})|}{2C\|\mathbf{t}\|^3},$$

so gilt

$$|r(s\mathbf{t})| \leq Cs^3\|\mathbf{t}\|^3 < \frac{1}{2}s^2q(\mathbf{t}) = \left|\frac{1}{2}q(s\mathbf{t})\right|,$$

also hat

$$f(R) - f(Q) = \frac{1}{2}q(s\mathbf{t}) + r(s\mathbf{t})$$

das selbe Vorzeichen wie $\frac{1}{2}q(s\mathbf{t})$, d. h. wie $q(\mathbf{t})$. Für R in einer Umgebung von P kommen somit bei dieser Differenz die selben Vorzeichen vor wie bei der quadratischen Form q. Hieraus folgt bereits die Behauptung (iii). Ist h definit, so gibt es laut Aufgabe 25 eine positive Zahl c, so dass $c|q(\mathbf{t})| \geq \|\mathbf{t}\|^2$ für alle $\mathbf{t}$, und dann gelten die obigen Abschätzungen für alle Punkte R mit der Eigenschaft $d(P,R) < \min(e, \frac{1}{2Cc^2})$. □

Der Graph einer nichtausgearteten quadratischen Form auf einem zweidimensionalen Vektorraum ist ein Hyperboloid, und zwar ein elliptisches im definiten Fall und ein hyperbolisches im indefiniten Fall. Da letzteres an einen Sattel erinnert, nennt man P im Fall (iii) einen Sattelpunkt von f (sogar für beliebige Dimension).

Es stellt sich wieder die Frage, wie man das zweite Differential praktisch berechnet, wenn auf dem affinen Raum A ein Koordinatensystem vorgegeben ist und f als Funktion von n Variablen $x_1, \ldots, x_n$ interpretiert wird.

Betrachten wir zunächst den Fall $n = 1$. Dann ist das (erste) Differential gegeben durch

$$l(t, x) = f'(x)t.$$

Das Differential dieser Funktion von x für festes t ist

$$h(t, u) = f''(x)tu,$$

wobei f'' die zweite Ableitung (also die Ableitung der Ableitung) bezeichnet. Die Spezialisierung ist dann $q(t) = f''(x)t^2$. Wir erinnern uns, dass Leibniz das Differential l mit df bezeichnet hat, was auf die Formel $df = f'dx$ führte. Für die Spezialisierung q des zweiten Differentials h schreibt man traditionell d^2f (gelesen d zwei f), so dass sich ergibt

$$d^2 f = f''\, dx^2.$$

Dies rechtfertigt halbwegs die Leibnizsche Schreibweise

$$\frac{d^2 f}{dx^2}$$

(gelesen „d zwei f nach d x [zum] Quadrat“) für die zweite Ableitung.

Nun kommen wir zum Fall von n Variablen. Das erste Differential hat die Form

$$l(t_1, \ldots, t_n, x_1, \ldots, x_n) = a_1(x_1, \ldots, x_n)t_1 + \ldots + a_n(x_1, \ldots, x_n)t_n.$$

Bilden wir hiervon das Differential für feste t_1, …, t_n, so erhalten wir

$$\begin{aligned} h(t_1, \ldots, t_n, u_1, \ldots, u_n) = {} & h_{11}t_1u_2 + h_{12}t_1u_2 + \ldots + h_{1n}t_1u_n \\ & + h_{21}t_2u_1 + h_{22}t_2u_2 + \ldots + h_{2n}t_2u_n \\ & \ldots \\ & + h_{n1}t_nu_1 + h_{n2}t_nu_2 + \ldots + h_{nn}t_nu_n, \end{aligned}$$

wobei h_{ij} die partielle Ableitung von a_i nach x_j ist. Da a_i selbst die partielle Ableitung von f nach x_i ist, handelt es sich um eine zweite partielle Ableitung, für die sich die Schreibweise

$$h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

(gelesen „d zwei f nach d x i d x j“) eingebürgert hat. Hier haben Zähler und Nenner für sich genommen allerdings keinen Sinn.

Interpretieren wir wieder die Variablen x_i als Funktionen, die ebenfalls Differentiale besitzen, so ergibt sich die klassische Schreibweise für das zweite Differential (genauer gesagt, seine Spezialisierung)

$$\begin{aligned} d^2 f = \frac{\partial^2 f}{\partial x_1^2} dx_1^2 + 2\frac{\partial^2 f}{\partial x_1 \partial x_2} dx_1 dx_2 + \ldots + 2\frac{\partial^2 f}{\partial x_1 \partial x_n} dx_1 dx_n \\ + \frac{\partial^2 f}{\partial x_2^2} dx_2^2 \qquad + \ldots + 2\frac{\partial^2 f}{\partial x_2 \partial x_n} dx_2 dx_n \\ \ddots \qquad \vdots \\ + \frac{\partial^2 f}{\partial x_n^2} dx_n^2. \end{aligned}$$

Beispiel. Wir betrachten wieder die Funktion

$$f(x_1, x_2) = 3\sqrt{x_1^2 + x_2^2 + 1} - x_1 - 2x_2$$

aus dem vorigen Absatz. Ihre zweiten partiellen Ableitungen sind gegeben durch

$$\begin{aligned} \frac{\partial^2 f}{\partial x_1^2} &= 3\frac{x_2^2 + 1}{(x_1^2 + x_2^2 + 1)^{3/2}} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} &= -3\frac{x_1 x_2}{(x_1^2 + x_2^2 + 1)^{3/2}} \\ \frac{\partial^2 f}{\partial x_2^2} &= 3\frac{x_1^2 + 1}{(x_1^2 + x_2^2 + 1)^{3/2}} \end{aligned}$$

Im stationären Punkt $\left(\frac{1}{2}, 1\right)$ haben sie die Werte

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{16}{9}, \qquad \frac{\partial^2 f}{\partial x_1 \partial x_2} = -\frac{4}{9}, \qquad \frac{\partial^2 f}{\partial x_2^2} = \frac{10}{9}.$$

Die Hesse-Form ist somit gleich

$$h(\mathbf{t}, \mathbf{u}) = \frac{16}{9} t_1 u_1 - \frac{4}{9} t_1 u_2 - \frac{4}{9} t_2 u_1 + \frac{10}{9} t_2 u_2$$

und ihre Spezialisierung gleich

$$\frac{16}{9} t_1^2 - \frac{8}{9} t_1 t_2 + \frac{10}{9} t_2^2.$$

Durch quadratische Ergänzung wird letztere zu

$$\frac{16}{9}(t_1 - \frac{1}{4} t_2)^2 + t_2^2.$$

Diese quadratische Form ist positiv definit, also hat f an der Stelle $\left(\frac{1}{2}, 1\right)$ nach Satz 12 ein lokales Minimum.

11 Matrizen

Das Rechnen mit linearen Substitutionen und Bilinearformen bringt viel Schreibarbeit mit sich. Die wesentliche Information über eine lineare Substutution, z. B.

$$\begin{aligned} u_1 &= x_1 + 3x_2 + 4x_3 \\ u_2 &= x_1 + 4x_2 + 5x_3 \\ u_3 &= x_1 + \ x_2 + \ x_3 \end{aligned}$$

(vgl. S. 12) ist in dem Schema von Zahlen

$$\begin{pmatrix} 1 & 3 & 4 \\ 1 & 4 & 5 \\ 1 & 1 & 1 \end{pmatrix}$$

enthalten, das man eine Matrix nennt. Verkettet man die obige lineare Substutution mit einer weiteren, z. B.

$$\begin{aligned} x_1 &= t_1 + 2t_2 \\ x_2 &= t_1 \\ x_3 &= t_1 - 3t_2 \end{aligned}$$

deren Matrix

$$\begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 1 & -3 \end{pmatrix}$$

ist, so ergibt sich die lineare Substitution

$$\begin{aligned} u_1 &= \ 8t_1 - 10t_2 \\ u_2 &= 10t_1 - 13t_2 \\ u_3 &= \ 3t_1 - \ \ t_2 \end{aligned}$$

deren Matrix man das Produkt der gegebenen Matrizen nennt:

$$\begin{pmatrix} 1 & 3 & 4 \\ 1 & 4 & 5 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 0 \\ 1 & -3 \end{pmatrix} = \begin{pmatrix} 8 & -10 \\ 10 & -13 \\ 3 & -1 \end{pmatrix}$$

Anstelle von Zahlen könnten hier auch Elemente eines Ringes stehen. Wir werden Matrizen mit fettgedruckten Großbuchstaben bezeichnen und den Eintrag einer Matrix $\mathbf{A}$, der in der $i$ten Zeile und $j$ten Spalte steht, mit a_{ij}.

Definition 22. *Das Produkt von Matrizen* $\mathbf{A}$ *und* $\mathbf{B}$ *mit Einträgen in einem Ring* R *ist definiert, wenn* $\mathbf{A}$ *genau so viele Spalten hat, wie* $\mathbf{B}$ *Zeilen hat (sagen wir* n*). Die Matrix* $\mathbf{C} = \mathbf{AB}$ *hat dann genau so vielen Zeilen wie* $\mathbf{A}$ *und genau so vielen Spalten wie* $\mathbf{B}$*, und ihre Einträge sind*

$$c_{ij} = a_{i1}b_{1j} + \ldots + a_{in}b_{nj}.$$

Die Multiplikation von Matrizen ist assoziativ, aber nicht kommutativ. Der trivialen Substitution $u_1 = x_1, \ldots, u_n = x_n$ entspricht die Einheitsmatrix $\mathbf{E}$ (oder $\mathbf{E}_n$, wenn man die Anzahl der Zeilen und Spalten angeben will), bei der auf der Diagonalen Einsen und an allen anderen Stellen Nullen stehen. Für alle Matrizen gilt $\mathbf{AE} = \mathbf{A}$, $\mathbf{EB} = \mathbf{B}$, falls die Produkte definiert sind.

Eine Matrix, die nur eine Zeile (bzw. nur eine Spalte) hat, bezeichnet man als Zeilenvektor (bzw. Spaltenvektor). Man kann eine lineare Substitution auch durch Matrizenmultiplikation ausdrücken, z. B. die eingangs angeführte als

$$\begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 4 \\ 1 & 4 & 5 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Deshalb ordnet man Koordinaten bevorzugt in Spaltenvektoren an. Zeilenvektoren erscheinen hingegen, wenn man eine Linearform durch Matrizenmultiplikation ausdrückt:

$$l(x_1, \ldots, x_n) = a_1x_1 + \ldots + a_nx_n = \begin{pmatrix} a_1 & \ldots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ v_n \end{pmatrix}.$$

Auch bei einer Bilinearform, z. B.

$$\begin{aligned} b(x_1, x_2, y_1, y_2, y_3) = {} & 3x_1y_1 - 2x_1y_2 + 4x_1y_3 \\ & + x_2y_1 + 5x_2y_2 - x_2y_3 \end{aligned}$$

(vgl. S. 9) kann man die wesentliche Information platzsparend in einer Matrix wiedergeben, im gegebenen Fall

$$\mathbf{B} = \begin{pmatrix} 3 & -2 & 4 \\ 1 & 5 & -1 \end{pmatrix}$$

Man nennt sie die Gramsche Matrix der Bilinearform b. Vertauscht man die Buchstaben x und y, so entsteht wieder eine Bilinearform

$$c(x_1, x_2, x_3, y_1, y_2) = b(y_1, y_2, x_1, x_2, x_3).$$

Ihre Gramsche Matrix hat die Einträge $c_{ij} = b_{ji}$. Man nennt sie die transponierte Matrix $\mathbf{B}^\top$. Natürlich ist eine Bilinearform genau dann symmetrisch, wenn ihre Gramsche Matrix symmetrisch ist in dem Sinne, dass $\mathbf{B} = \mathbf{B}^\top$ gilt. Man kann sich leicht klar machen, dass

$$(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top.$$

Auch den Wert einer Bilinearform kann man durch Matrizenmultiplikation ausdrücken, im obigen Beispiel etwa

$$b(x_1, x_2, y_1, y_2, y_3) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 3 & -2 & 4 \\ 1 & 5 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

Bezeichnet man die aus den Variablen gebildeten Spaltenvektoren mit $\mathbf{x}$ bzw. $\mathbf{y}$, so ist der Wert einer Bilinearform mit Gramscher Matrix $\mathbf{B}$ gleich

$$\mathbf{x}^\top \mathbf{B} \mathbf{y}.$$

Ist die Anzahl der x- und y-Variablen gleich, so ist der Wert der Spezialisierung gegeben durch

$$\mathbf{x}^\top \mathbf{B} \mathbf{x}.$$

Wir wissen, dass jede quadratische Form mit Koeffizienten in einem Ring R durch Spezialisierung aus einer Bilinearform entsteht, und nach Folgerung 1 sogar aus einer eindeutig bestimmten symmetrischen Bilinearform, wenn 2 in R invertierbar ist. Man kann dann von der Gramschen Matrix einer quadratischen Form sprechen. Nehmen wir eine Substitution $\mathbf{x} = \mathbf{Au}$ vor, so ist der Wert der quadratischen Form in den Variablen $u_1, \ldots, u_n$ wegen der Assoziativität der Matrizenmultiplikation gleich

$$(\mathbf{Au})^\top \mathbf{B} (\mathbf{Au}) = \mathbf{u}^\top (\mathbf{A}^\top \mathbf{BA}) \mathbf{u}.$$

Die Gramsche Matrix in den neuen Variablen ist also die symmetrische Matrix

$$\mathbf{A}^\top \mathbf{BA}.$$

Eine solche Matrix nennt man zur Matrix $\mathbf{B}$ kongruent (vorausgesetzt, $\mathbf{A}$ ist regulär). Eine Quadratische Form hat genau dann Diagonalform, wenn ihre Gramsche Matrix im offensichtlichen Sinne eine Diagonalmatrix ist.

Als Anwendung des Begriffs der Bilinearform hatten wir die Hessesche Form einer Funktion f an einer Stelle des Definitionsbereiches betrachtet. Wählen wir ein Koordinatensystem $x_1, \ldots, x_n$, so hat die Hesse-Form eine Gramsche Matrix $\mathbf{H}$, genannt Hesse-Matrix, mit den Einträgen

$$h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

12 Lineare Gleichungssysteme

Wir betrachten lineare[7] Gleichungssysteme, z. B.

$$\begin{aligned} x_1 + 3x_2 + 4x_3 &= 2 \\ x_1 + 4x_2 + 5x_3 &= 1 \\ x_1 + \ x_2 + \ x_3 &= 1 \end{aligned}$$

In der Schule werden solche Gleichungssysteme meist durch Elimination gelöst. Dabei ist es nicht leicht, die Übersicht darüber zu behalten, welche Gleichungen noch gebraucht werden. Statt dessen kann man das System in äquivalente Systeme umformen und dabei schrittweise vereinfachen.

Um x_1 aus der zweiten und dritten Gleichung zu eliminieren, subtrahieren wir die erste Gleichung von jenen Gleichungen:

$$\begin{aligned} x_1 + 3x_2 + 4x_3 &= 2 \\ x_2 + \ x_3 &= -1 \\ -2x_2 - 3x_3 &= -1 \end{aligned}$$

Nun eliminieren wir x_2 aus der dritten Gleichung, indem wir zu ihr das Doppelte der zweiten Gleichung addieren:

$$\begin{aligned} x_1 + 3x_2 + 4x_3 &= 2 \\ x_2 + \ x_3 &= -1 \\ -x_3 &= -3 \end{aligned}$$

Wir haben hier das Gaußsche Eliminationsverfahren angewendet, das nach Carl Friedrich Gauß (1777–1855) benannt ist und das System im Idealfall in eine Dreiecksform bringt.

Wir können die Lösung für x_3 aus der letzten Gleichung ablesen und in die anderen einsetzen. Wir bleiben aber bei unserer Methode und addieren die letzte Gleichung zur zweiten sowie das Vierfache der letzten Gleichung zur ersten:

$$\begin{aligned} x_1 + 3x_2 &= -10 \\ x_2 &= -4 \\ -x_3 &= -3 \end{aligned}$$

[7] Wenn man alle Terme auf die linken Seiten bringt, entstehen genaugenommen nur dann Linearformen, wenn die rechten Seiten vorher gleich Null waren. In diesem Fall spricht man von homogenen Gleichungen.

Nun subtrahieren wir das Dreifache der zweiten Gleichung von der ersten und multiplizieren die letzte Gleichung mit -1:

$$\begin{aligned} x_1 \qquad\quad &= 2 \\ x_2 \qquad &= -4 \\ x_3 &= 3 \end{aligned}$$

Da alle Operationen umkehrbar waren, ist das vereinfachte Gleichungssystem äquivalent zu dem Ausgangssystem. Die beschriebene Methode nennt man nach Gauß und Wilhelm Jordan (1842–1899) und mitunter den Gauß-Jordan-Algorithmus.

Auch hier kann man sich durch Verwendung von Matrizen viel Schreibarbeit sparen. Das Gleichungssystem lässt sich in der Form $\mathbf{Ax} = \mathbf{b}$ schreiben, wobei $\mathbf{A}$ die sogenannte Koeffizientenmatrix des Systems und $\mathbf{b}$ ein Spaltenvektor von Konstanten ist. Um die die gesamte Information aufzuzeichnen, benötigen wir die erweiterte Koeffizientenmatrix, die neben $\mathbf{A}$ den Spaltenvektor $\mathbf{b}$ enthält:

$$\begin{pmatrix} 1 & 3 & 4 & 2 \\ 1 & 4 & 5 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

Unsere Umformungen erscheinen nun als Zeilenoperationen. Wir können

- ein Vielfaches einer Zeile zu einer anderen Zeile addieren und
- eine Zeile mit einer von Null verschiedenen Zahl multiplizieren.

Die Folge der obigen Umformungen lässt sich nun wie folgt notieren:

$$\begin{pmatrix} 1 & 3 & 4 & 2 \\ 1 & 4 & 5 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 3 & 4 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & -2 & -3 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 3 & 4 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & -1 & -3 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 3 & 0 & -10 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & -1 & -3 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & -1 & -3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 3 \end{pmatrix}.$$

Wir können mit dieser Methode nicht nur Gleichungssysteme lösen, sondern auch lineare Substitutionen umkehren, falls diese umkehrbar sind. Betrachten wir wieder unser Beispiel

$$\begin{aligned} x_1 + 3x_2 + 4x_3 &= u_1 \\ x_1 + 4x_2 + 5x_3 &= \qquad u_2 \\ x_1 + \ x_2 + \ x_3 &= \qquad\quad u_3 \end{aligned}$$

Diesmal gehen wir sofort zur Matrizenschreibweise über und führen die selben umkehrbaren Zeilenoperationen wie oben aus, die man sich natürlich als Operationen auf dem Gleichungssystem vorstellen kann:

$$\begin{pmatrix}1&3&4&1&0&0\\1&4&5&0&1&0\\1&1&1&0&0&1\end{pmatrix}, \begin{pmatrix}1&3&4&1&0&0\\0&1&1&-1&1&0\\0&-2&-3&-1&0&1\end{pmatrix}, \begin{pmatrix}1&3&4&1&0&0\\0&1&1&-1&1&0\\0&0&-1&-3&2&1\end{pmatrix},$$
$$\begin{pmatrix}1&3&0&-11&8&4\\0&1&0&-4&3&1\\0&0&-1&-3&2&1\end{pmatrix}, \begin{pmatrix}1&0&0&1&-1&1\\0&1&0&-4&3&1\\0&0&-1&-3&2&1\end{pmatrix}, \begin{pmatrix}1&0&0&1&-1&1\\0&1&0&-4&3&1\\0&0&1&3&-2&-1\end{pmatrix}.$$

Damit haben wir unser Gleichungssystem in das äquivalente System

$$\begin{aligned} x_1 &= u_1 - u_2 + u_3,\\ x_2 &= -4u_1 + 3u_2 + u_3,\\ x_3 &= 3u_1 - 2u_2 - u_3. \end{aligned}$$

umgewandelt. In dem Beispiel am Anfang von Abschnitt 4 war dieses Ergebnis ohne Rechenweg mitgeteilt (und nachgeprüft) worden.

Man kann das vorangehende Ergebnis auch so interpretieren, dass wir nacheinander zwei lineare Substitutionen vornehmen und im Ergebnis die triviale Substitution erhalten. Dabei kommt es nicht einmal auf die Reihenfolge an. In der Sprache der Matrizenmultiplikation bedeutet dies

$$\begin{pmatrix}1&3&4\\1&4&5\\1&1&1\end{pmatrix}\begin{pmatrix}1&-1&1\\-4&3&1\\3&-2&-1\end{pmatrix} = \begin{pmatrix}1&-1&1\\-4&3&1\\3&-2&-1\end{pmatrix}\begin{pmatrix}1&3&4\\1&4&5\\1&1&1\end{pmatrix} = \begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}.$$

Definition 23. *Eine Matrix, die bei Multiplikation mit einer gegebenen Matrix* $\mathbf{A}$ *die Einheitsmatrix ergibt, heißt inverse Matrix von* $\mathbf{A}$*. Wenn* $\mathbf{A}$ *eine inverse Matrix besitzt, so nennt man* $\mathbf{A}$ *eine reguläre Matrix, andernfalls eine singuläre Matrix.*

Sind $\mathbf{B}$ und $\mathbf{C}$ inverse Matrizen von $\mathbf{A}$, so ist $\mathbf{B} = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{C}$. Die inverse Matrix ist also eindeutig bestimmt und wird mit $\mathbf{A}^{-1}$ bezeichnet.

Der Gauß-Jordan-Algorithmus ist auf lineare Gleichungssysteme mit Koeffizienten in einem beliebigen Körper anwendbar. Allerdings führt er mitunter nicht auf eine Diagonalform, so z. B. bei folgendem Gleichungssystem.

$$\begin{aligned} 2x_1 - 3x_2 + 2x_3 + x_4 &= 4\\ 4x_1 - x_2 - 6x_3 + 6x_4 &= 5\\ 2x_1 - x_2 - 2x_3 + 2x_4 &= 3\\ x_1 + x_2 - 4x_3 + x_4 &= 1 \end{aligned}$$

In Matrizenschreibweise beginnt der Algorithmus mit folgenden Schritten:

$$\begin{pmatrix} 2 & -3 & 2 & 1 & 4 \\ 4 & -1 & -6 & 6 & 5 \\ 2 & -1 & -2 & 2 & 3 \\ 1 & 1 & -4 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & -3 & 2 & 1 & 4 \\ 0 & 5 & -10 & 4 & -3 \\ 0 & 2 & -4 & 1 & -1 \\ 0 & \frac{5}{2} & -5 & \frac{1}{2} & -1 \end{pmatrix} \begin{pmatrix} 2 & -3 & 2 & 1 & 4 \\ 0 & 5 & -10 & 4 & -3 \\ 0 & 0 & 0 & -\frac{3}{5} & \frac{1}{5} \\ 0 & 0 & 0 & -\frac{3}{2} & \frac{1}{2} \end{pmatrix}$$

In der dritten Zeile kommt erst an der vierten Stelle ein von Null verschiedenes Element. Wäre das in der vierten Zeile nicht so, dann müsste man die dritte und vierte Zeile vertauschen und wie gewöhnlich fortfahren. In der vorliegenden Situation aber ignoriert man die dritte Spalte und eliminiert den untersten Eintrag in der vierten Spalte. Dann multipliziert man jede Spalte mit dem Inversen des ersten von Null verschiedenen Eintrags:

$$\begin{pmatrix} 2 & -3 & 2 & 1 & 4 \\ 0 & 5 & -10 & 4 & -3 \\ 0 & 0 & 0 & -\frac{3}{5} & \frac{1}{5} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & -\frac{3}{2} & 1 & \frac{1}{2} & 2 \\ 0 & 1 & -2 & \frac{4}{5} & -\frac{3}{5} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

In der letzten Zeile stehen nur noch Nullen, und die zugehörige Gleichung ist immer erfüllt. Hätten wir auf der rechten Seite eine von Null verschiedene Zahl erhalten, so wäre diese Gleichung unerfüllbar, und die Lösungsmenge des Systems wäre leer.

Man könnte die Nummerierung der Variablen ändern und dadurch die zweite und dritte Spalte vertauschen. Dann hätte man eine Dreiecksform, in der auch Nullen auf der Diagonalen vorkommen.

Nun beseitigen wir die Einträge oberhalb der Diagonalen, soweit möglich, durch weitere Zeilenoperationen:

$$\begin{pmatrix} 1 & -\frac{3}{2} & 1 & 0 & \frac{13}{6} \\ 0 & 1 & -2 & 0 & -\frac{1}{3} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & -2 & 0 & \frac{5}{3} \\ 0 & 1 & -2 & 0 & -\frac{1}{3} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Unser Gleichungssystem ist also äquivalent zu dem System

$$\begin{aligned} x_1 \qquad -2x_3 \qquad &= \frac{5}{3} \\ x_2 - 2x_3 \qquad &= -\frac{1}{3} \\ x_4 &= -\frac{1}{3} \end{aligned}$$

Wählen wir für die Variable x_3 einen beliebigen Wert, beispielsweise 7, so sind die Werte der anderen Variablen eindeutig bestimmt, nämlich

$$x_1 = \frac{5}{3} + 2 \cdot 7, \quad x_2 = -\frac{1}{3} + 2 \cdot 7, \quad x_3 = 7, \quad x_4 = -\frac{1}{3}.$$

Man nennt x_3 darum eine freie Variable und die anderen gebundene Variablen. Die Lösungsmenge ist die unendliche Menge

$$\{(2t + \tfrac{5}{3}, 2t - \tfrac{1}{3}, t, -\tfrac{1}{3}) \mid t \in \mathbf{Q}\},$$

falls wir nach rationalen Lösungen suchen. Vor Einführung der Mengenschreibweise sagte man, die allgemeine Lösung sei von der Form

$$x_1 = 2t + \frac{5}{3}, \quad x_2 = 2t - \frac{1}{3}, \quad x_3 = t, \quad x_4 = -\frac{1}{3}$$

und nannte t einen Parameter.

Die angegebene Darstellung und die Wahl von freien Variablen ist allerdings nicht eindeutig. Mit einer umkehrbaren Substitution, z. B. $t = 3s + 1$, erhalten wir

$$x_1 = 6s + \frac{11}{3}, \quad x_2 = 6s + \frac{5}{3}, \quad x_3 = 3s + 1, \quad x_4 = -\frac{1}{3}.$$

Es kann auch mehrere freie Variablen geben. So hat z. B. das System, das aus der einzigen Gleichung

$$3x_1 - 5x_2 + 2x_3 = 3$$

besteht, die allgemeine Lösung

$$x_1 = t_1 + 2t_2 + 1, \quad x_2 = t_1, \quad x_3 = t_1 - 3t_2,$$

die von zwei Parametern abhängt, vgl. Aufgabe 23.

Satz 13. *Es sei* $\mathbf{A}$ *eine Matrix. Das lineare Gleichungssystem* $\mathbf{Ax} = \mathbf{b}$ *ist genau dann für einen beliebigen Spaltenvektor* $\mathbf{b}$ *eindeutig lösbar, wenn* $\mathbf{A}$ *regulär ist, und in diesem Fall hat* $\mathbf{A}$ *genauso viele Zeilen wie Spalten.*

Beweis. Die Matrix $\mathbf{A}$ habe m Zeilen und n Spalten. Wenn wir die Variablen $x_1, \ldots, x_n$ geeignet nummerieren, so ist das Gleichungssystem äquivalent zu einem System, dessen Koeffizientenmatrix folgende Form hat:

$$\begin{pmatrix} 1 & 0 & \ldots & 0 & a_{1,k+1} & \ldots & a_{1,n} \\ 0 & 1 & \ldots & 0 & a_{2,k+1} & \ldots & a_{2,n} \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 1 & a_{k,k+1} & \ldots & a_{k,n} \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{pmatrix} \tag{8}$$

Dies folgt durch Anwendung des Gauß-Jordan-Algorithmus. Die linke obere Untermatrix der Größe $k \times k$ ist nun eine Einheitsmatrix. Ist $k < m$, so hat das System für $b_m \neq 0$ keine Lösung. Ist $k = m$, so hat die allgemeine Lösung die Form

$$\begin{aligned}
x_1 &= b_1 - a_{1,m+1}t_1 - \ldots - a_{1,n}t_{n-k} \\
&\vdots \\
x_k &= b_k - a_{k,k+1}\,t_1 - \ldots - a_{k,n}t_{n-k} \\
x_{k+1} &= \qquad\quad t_1 \\
&\vdots \qquad\qquad\quad \ddots \\
x_n &= \qquad\qquad\qquad\quad t_{n-k}
\end{aligned}$$

Die Lösung ist nur eindeutig, wenn $k = n$ ist. In diesem Fall gilt also $m = n$, der Gauß-Jordan-Algorithmus verwandelt $\mathbf{A}$ in die Einheitsmatrix, und wenn man ihn auf die Matrix $(\mathbf{A}, \mathbf{E})$ anwendet, liefert er $(\mathbf{E}, \mathbf{A}^{-1})$.

Ist umgekehrt $\mathbf{A}$ regulär, so ist das Gleichungssystem $\mathbf{Ax} = \mathbf{b}$ äquivalent zu dem System $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, welches für beliebige $\mathbf{b}$ eindeutig lösbar ist. □

13 Unterräume und Basen

Am Ende von Abschnitt 7 haben wir im Kleingedruckten bereits den Begriff des affinen Raums und des Unterraums eingeführt, wobei nur ein anschauliches Verständnis verlangt war. Nun wollen wir dieses Verständnis vertiefen.

Definition 24. *Eine Teilmenge W eines Vektorraums V über einem Körper K heißt (linearer) Unterraum, wenn sie abgeschlossen unter den Operationen der Addition und der Skalarmultiplikation mit Elementen von K ist. Eine Teilmenge A von V heißt affiner Unterraum, wenn es einen linearen Unterraum W gibt, so dass für jedes Element $\mathbf{a}$ von A gilt*

$$A = \{\mathbf{a} + \mathbf{t} \mid \mathbf{t} \in W\}.$$

Gilt die letzte Gleichung für ein Element $\mathbf{a}$ einer Teilmenge A, so gilt sie für alle. Ein linearer Unterraum eines Vektorraums ist natürlich selbst ein Vektorraum, ein affiner Unterraum im allgemeinen nicht.

Satz 14. *Die Lösungsmenge eines linearen Gleichungssystems mit m Gleichungen in n Variablen mit Koeffizienten in einem Körper K ist leer oder ein affiner Unterraum von K^n der Dimension mindestens $n - m$. Ist das System homogen, so ist die Lösungsmenge ein linearer Unterraum.*

Beweis. Ist die Lösungsmenge nicht leer, so ist die allgemeine Lösung, wie wir im Beweis von Satz 13 gesehen haben, gleich

$$\begin{pmatrix} x_1 \\ \vdots \\ x_k \\ x_{k+1} \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} - t_1 \begin{pmatrix} a_{1,k+1} \\ \vdots \\ a_{k,n+1} \\ 1 \\ \vdots \\ 0 \end{pmatrix} - t_{n-k} \begin{pmatrix} a_{n,k+1} \\ \vdots \\ a_{n,n+1} \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

wobei a_{ij} und b_i die Koeffizienten des umgeformten Gleichungssystems bezeichnen. Natürlich muss b_i für $i > k$ gleich Null sein, damit eine Lösung existiert.

Ist nun $b_i = 0$ für alle i, so hat die allgemeine Lösung die Form

$$\mathbf{x} = t_1\mathbf{w}_1 + \ldots + t_{n-k}\mathbf{w}_{n-k},$$

wobei sich die Spaltenvektoren $\mathbf{w}_i$ aus der obigen Formel ablesen lassen und $t_i \in K$ beliebig ist. Die Menge solcher Vektoren bildet offensichtlich einen Unterraum W von K^n. Die Vektoren $\mathbf{w}_1, \ldots, \mathbf{w}_{n-k}$ bilden eine Basis von W, denn die Koordinaten t_i eines Vektors $\mathbf{x}$ bezüglich dieser Basis lassen sich als $t_i = x_{k+i}$ ablesen und sind somit eindeutig durch $\mathbf{x}$ bestimmt. Insbesondere ist die Dimension von W gleich $n - k \geq n - m$.

Sind hingegen $b_1 \ \ldots, b_k$ beliebig, so ist jeder Lösungsvektor von der Form

$$\mathbf{x} = \mathbf{a} + t_1\mathbf{w}_1 + \ldots + t_{n-k}\mathbf{w}_{n-k}, \tag{9}$$

wobei $\mathbf{a}$ gegeben ist durch $a_i = b_i$ für $i \leq k$ und $a_i = 0$ für $i > k$. Diese Vektoren bilden laut Definition einen affinen Unterraum. □

Folgerung 3. *Sind $l_1, \ldots, l_m$ Linearformen auf einem Vektorraum der Dimension n, so ist die Teilmenge der Vektoren $\mathbf{x}$ mit den Eigenschaften*

$$l_1(\mathbf{x}) = 0, \quad \ldots, \quad l_m(\mathbf{x}) = 0$$

ein Unterraum der Dimension mindestens $n - m$.

Beweis. Jeder Vektor lässt sich bezüglich einer fest gewählten Basis in der Form

$$\mathbf{x} = x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n$$

darstellen, und dann ist

$$l_i(\mathbf{x}) = a_{i1}x_1 + \ldots + a_{in}x_n.$$

Die Bedingungen $l_1(\mathbf{x}) = 0, \ldots, l_m(\mathbf{x}) = 0$ lassen sich nun als homogenes lineares Gleichungssystem schreiben, deren Koeffizientenmatrix die Einträge a_{ij} hat. Die Spaltenvektoren $\mathbf{w}_i$ aus dem Beweis von Satz 14 sind die Koordinatenvektoren gewisser Vektoren von V. Nennen wir diese ebenfalls $\mathbf{w}_i$, so besteht die gesuchte Menge aus allen Vektoren der Form (9), wobei $t_i \in K$, ist also ein Unterraum von V der Dimension $n - k$. □

Sind U und W Unterräume eines Vektorraums V, so ist zwar der Durchschnitt $U \cap W$ ein Unterraum, die Vereinigung $U \cup W$ aber im Allgemeinen nicht. Statt dessen ist die Summe der beiden Unterräume

$$U + W = \{\mathbf{u} + \mathbf{w} \mid \mathbf{u} \in U, \mathbf{w} \in W\}$$

wieder ein Unterraum.

Definition 25. *Wir sagen, der Vektorraum V sei die direkte Summe der Unterräume U und W, wenn sich jeder Vektor $\mathbf{x} \in V$ in der Form $\mathbf{x} = \mathbf{u} + \mathbf{w}$ schreiben lässt, wobei $\mathbf{u} \in U$ und $\mathbf{w} \in W$ eindeutig bestimmt sind.*

Lemma 4. *Ist $U + W = V$ und $U \cap W = \{\mathbf{0}\}$, so ist V die direkte Summe von U und W, und umgekehrt.*

Beweis. Die erste Bedingung besagt, dass sich jedes $\mathbf{x} \in V$ in der Form $\mathbf{x} = \mathbf{u} + \mathbf{w}$ schreiben lässt, wobei $\mathbf{u} \in U$ und $\mathbf{w} \in W$. Gibt es eine weitere solche Darstellung $\mathbf{x} = \mathbf{u}' + \mathbf{w}'$, so ist $\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w}$ ein Element sowohl von U als auch von W, also von $U \cap W$. Ist dieser Durchschnitt gleich $\mathbf{0}$, so folgt $\mathbf{u} = \mathbf{u}'$ und $\mathbf{w} = \mathbf{w}'$.

Umgekehrt sei V die direkte Summe von U und W. Jedes Element $\mathbf{x}$ von $U \cap W$ lässt sich in der Form $\mathbf{x} + \mathbf{0}$ mit $\mathbf{x} \in U$ und $\mathbf{0} \in W$ und auch in der Form $\mathbf{0} + \mathbf{x}$ mit $\mathbf{0} \in U$ und $\mathbf{x} \in W$ schreiben. Aus der Eindeutigkeit folgt $\mathbf{x} = \mathbf{0}$. □

Wir müssen auch den Begriff der Basis eines Vektorraums V vertiefen. Sie ist durch zwei Eigenschaften charakterisiert. Nun betrachten wir jede dieser Eigenschaften getrennt.

Definition 26. *Es sei V ein Vektorraum über einem Körper K.*

(i) Ein Vektor $\mathbf{x} \in V$ heißt linear abhängig von den Vektoren $\mathbf{v}_1, \ldots, \mathbf{v}_n$ von V, wenn er sich als Linearkombination

$$\mathbf{x} = x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n$$

mit Elementen $x_1, \ldots, x_n \in K$ darstellen lässt. Der von den Vektoren $\mathbf{v}_1, \ldots, \mathbf{v}_n$ erzeugte Unterraum ist die Menge aller Vektoren, die von $\mathbf{v}_1, \ldots, \mathbf{v}_n$ linear abhängig sind.

(ii) Eine Menge von Vektoren $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eines Vektorraums heißt linear unabhängig, wenn für beliebige Elemente $x_1, \ldots, x_n$ des Körpers K mit der Eigenschaft

$$x_1\mathbf{v}_1 + \ldots + x_n\mathbf{v}_n = \mathbf{0}$$

gilt, dass $x_1 = 0, \ldots, x_n = 0$.

Lemma 5. *Eine Menge von Vektoren in einem Vektorraum ist genau dann eine Basis, wenn sie linear unabhängig ist und den Vektorraum erzeugt.*

Beweis. Erzeugen $\mathbf{v}_1, \ldots, \mathbf{v}_n$ den Vektorraum V, so lässt sich jeder Vektor $\mathbf{x}$ als Linearkombination wie in (i) darstellen. Angenommen, es gibt eine weitere Darstellung

$$\mathbf{x} = x_1'\mathbf{v}_1 + \ldots + x_n'\mathbf{v}_n,$$

wobei $x_i' \in K$. Dann folgt

$$(x_1 - x_1')\mathbf{v}_1 + \ldots + (x_n - x_n')\mathbf{v}_n = \mathbf{0}.$$

Sind die gegebenen Vektoren zusätzlich linear unabhängig, so folgt $x_i - x_i' = 0$, also ist die Darstellung eindeutig.

Umgekehrt sei $\mathbf{v}_1, \ldots, \mathbf{v}_n$ eine Basis. Nach Definition erzeugt sie den Vektorraum, und die Koeffizienten in der Darstellung jedes Vektors als Linearkombination dieser Vektoren sind eindeutig bestimmt. Dies gilt inbesondere für den Nullvektor, und wegen

$$0\mathbf{v}_1 + \ldots + 0\mathbf{v}_n = \mathbf{0}$$

folgt die lineare Unabhängigkeit. □

Satz 15. *Jede linear unabhängige Teilmenge eines endlich erzeugten Vektorraums lässt sich zu einer Basis ergänzen.*

Da die leere Menge linear unabhängig ist, besitzt folglich jeder endlich erzeugte Vektorraum eine Basis.

Beweis. Angenommen, die Elemente $\mathbf{v}_1, \ldots, \mathbf{v}_k$ sind linear unabhängig. Da V endlich erzeugt ist, brauchen wir nur endlich viele Vektoren $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_l$ hinzuzufügen, so dass V von $\mathbf{v}_1, \ldots, \mathbf{v}_l$ erzeugt ist. Ist $k = l$, so sind wir wegen Lemma 5 fertig. Nun betrachten wir den Fall $k < l$.

Ist der Vektor $\mathbf{v}_{k+1}$ linear abhängig von $\mathbf{v}_1, \ldots, \mathbf{v}_k$, so lässt er sich in jeder Linearkombination von $\mathbf{v}_1, \ldots, \mathbf{v}_l$ durch die übrigen Vektoren ausdrücken. Letztere erzeugen also immer noch den Vektorraum V.

Ist hingegen $\mathbf{v}_{k+1}$ nicht linear abhängig von $\mathbf{v}_1, \ldots, \mathbf{v}_k$, so behaupten wir, dass die Elemente $\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}$ linear unabhängig sind. Ist nämlich

$$x_1\mathbf{v}_1 + \ldots + x_{k+1}\mathbf{v}_{k+1} = \mathbf{0}$$

für irgend welche $x_i \in K$, so muss $x_{k+1} = 0$ sein, denn sonst wäre

$$\mathbf{v}_{k+1} = -\frac{x_1}{x_{k+1}}\mathbf{v}_1 - \ldots - \frac{x_k}{x_{k+1}}\mathbf{v}_k,$$

also $\mathbf{v}_{k+1}$ linear abhängig von $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Wegen der linearen Unabhängigkeit dieser Vektoren müssen dann aber auch die übrigen Koeffizienten x_i gleich Null sein.

Wir haben in beiden Fällen die Differenz $l - k$ um eins verringert. Fahren wir auf diese Weise fort, erhalten wir schließlich eine Basis. □

Drücken wir die Vektoren einer Basis wie im Anschluss an Definition 9 als Linearkombination der Vektoren einer anderen Basis aus, so bilden die Koeffizienten eine Matrix $\mathbf{A}$, genannt Basiswechselmatrix. Vertauschen wir die Rollen der beiden Basen, so erhalten wir eine weitere Basiswechselmatrix $\mathbf{B}$. Führen wir erst den einen und dann den anderen Basiswechsel durch, so wird das Ergebnis duch das Produkt dieser Matrizen beschrieben. Wegen der Eindeutigkeit der Koordinaten folgt $\mathbf{AB} = \mathbf{E}$, $\mathbf{BA} = \mathbf{E}$. Die beiden Basiswechselmatrizen sind also regulär. Aus Satz 13 erhalten wir nun die schon erwähnte Tatsache, die erst die Definition der Dimension rechtfertigt:

Folgerung 4. *In einem endlich erzeugten Vektorraum hat jede Basis die gleiche Anzahl von Elementen. Die maximale Anzahl von linear unabhängigen Vektoren ist gleich der Dimension.*

Man kürzt die Dimension von V mit $\dim V$ ab. Statt „endlich erzeugt“ sagt man im Fall von Vektorräumen meist „endlichdimensional“. Nur solche betrachten wir hier.

Folgerung 5. *Ist W ein Unterraum von V, so ist $\dim W \leq \dim V$. Gilt $\dim V = \dim W$, so ist $V = W$.*

Eine Basis von W ist nämlich auch in V linear unabhängig, lässt sich also nach Satz 15 zu einer Basis von V ergänzen. Bei Gleichheit der Dimensionen muss sie schon selbst eine Basis von V sein.

Folgerung 6. *Ist V die direkte Summe der Unterräume U und W, so gilt $\dim V = \dim U + \dim W$.*

Vereinigen wir nämlich eine Basis von U mit einer Basis von W, so erhalten wir eine Basis von W, was direkt aus den Definitionen folgt.

Noch ein paae Worte dazu, wie man eine Basis eines Unterraums W praktisch bestimmen kann. Ist W wie in Folgerung 3 gegeben, so läuft es auf die Lösung eines Gleichungssystems hinaus. Alternativ kann W gegeben sein als der von gewissen Vektoren $\mathbf{v}_1, \ldots, \mathbf{v}_k$ erzeugte Unterraum. In diesem Fall kann man

- einen der Vektoren mit einem von Null verschiedenen Element von K multiplizieren,
- ein Vielfaches eines dieser Vektoren zu einem anderen addieren,
- zwei dieser Vektoren vertauschen,

ohne dass sich der von den Vektoren erzeugte Unterraum ändert. Ist z. B. $V = K^n$ der Raum der Spaltenvektoren und fügt man die Spaltenvektoren $\mathbf{v}_1, \ldots, \mathbf{v}_k$ zu einer Matrix $\mathbf{A}$ zusammen, so ändert sich also der von den Spalten von $\mathbf{A}$ erzeugte Unterraum W nicht, wenn man Spaltenoperationen vornimmt. Mit solchen Operationen (und der Umnummerierung von Koordinaten) lässt sich die Matrix auf eine Form bringen, die zu der in Gleichung (8) transponiert ist. Die nicht verschwindenden Spalten bilden dann eine Basis von W. Man kann nämlich die Koeffizienten einer Linearkombination von ihnen an den oberen Einträgen des entstehenden Spaltenvektors ablesen.

14 Orthogonalkomplemente

Der aus der Euklidischen Geometrie bekannte Begriff der Orthogonalität lässt sich durch das Skalarprodukt ausdrücken. Dessen positive Definitheit ist für viele Ergebnisse unerheblich.

Definition 27. *Es sei b eine symmetrische Bilinearform auf einem Vektorraum V.*

(i) Wir sagen, dass Vektoren $\mathbf{x}$ und $\mathbf{y}$ bezüglich b orthogonal sind, wenn $b(\mathbf{x}, \mathbf{y}) = 0$ ist.

(ii) Ein Vektor $\mathbf{x}$ heißt isotrop bezüglich b, wenn er zu sich selbst orthogonal ist, d. h. wenn $q(\mathbf{x}) = 0$ ist, wobei q die Spezialisierung von b bezeichnet.

(iii) Die Menge der Vektoren, die zu allen Vektoren eines Unterraums W orthogonal sind, nennt man das Orthogonalkomplement von W bezüglich b.

In Anlehnung an die Euklidische Geometrie schreibt man $\mathbf{x} \perp \mathbf{y}$, um auszudrücken, dass $\mathbf{x}$ orthogonal zu $\mathbf{y}$ ist, und bezeichnet das Orthogonalkomplement von W mit $W^\perp$. Leider geht die Abhängigkeit von der Wahl von b aus dieser Schreibweise nicht hervor. In Formeln kann man schreiben:

$$W^\perp = \{\mathbf{x} \in V \mid b(\mathbf{w}, \mathbf{x}) = 0 \text{ für alle } \mathbf{w} \in W\}.$$

Wenn W von $\mathbf{v}_1, \ldots, \mathbf{v}_k$ erzeugt wird, so gehört ein Vektor $\mathbf{x}$ genau dann zu $W^\perp$, wenn $b(\mathbf{v}_1, \mathbf{x}) = 0, \ldots, b(\mathbf{v}_k, \mathbf{x}) = 0$ ist. Dies folgt aus der Linearität von b bezüglich des ersten Arguments. Aus der Linearität von b bezüglich des zweiten Arguments folgt, dass $W^\perp$ ein Unterraum von V ist.

Direkt aus der Definition folgt, dass b genau dann ausgeartet ist, wenn der sogenannte Ausartungsraum $V^\perp$ ungleich $\{\mathbf{0}\}$ ist. Selbst für nicht ausgeartete Bilinearformen hat der Begriff der Orthogonalität ungewohnte Eigenschaften, wenn man keine Definitheit verlangt.

Beispiel. Auf dem Vektorraum K^2 der geordneten Paare aus dem Körper K ist die Bilinearform

$$b(\mathbf{x}, \mathbf{y}) = x_1 y_2 + x_2 y_1$$

symmetrisch und nicht ausgeartet. Alle Vektoren der Form $(t, 0)$ oder $(0, t)$ sind isotrop. Das Orthogonalkomplement des Unterraums $W = \{(t, 0) \mid t \in K\}$ ist wieder der Unterraum W.

Lemma 6. *Es sei b eine symmetrische Bilinearform auf V und W ein Unterraum von V. Dann ist $\dim W + \dim W^\perp \geq \dim V$. Ist außerdem b nicht ausgeartet, so gilt $\dim W + \dim W^\perp = \dim V$.*

Beweis. Wir wählen eine Basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$ von W und ergänzen sie gemäß Satz 15 zu einer Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ von V. Wir betrachten die Linearformen l_i auf V, die durch $l_i(\mathbf{x}) = b(\mathbf{v}_i, \mathbf{x})$ gegeben sind. Dann gehört $\mathbf{x}$ genau dann zu $W^\perp$, wenn $l_1(\mathbf{x}) = 0, \ldots, l_k(\mathbf{x}) = 0$. Nach Folgerung 3 ist $\dim W^\perp \geq n - k$.

Nun sei m_i die Einschränkung von l_i auf $W^\perp$. Ist $\mathbf{x} \in W^\perp$ und außerdem $m_{k+1}(\mathbf{x}) = 0, \ldots, m_n(\mathbf{x}) = 0$, so ist $\mathbf{x} \in V^\perp$. Ist also b nicht ausgeartet, so ist wegen Folgerung 3 nun $0 \geq \dim W^\perp - (n - k)$, also $\dim W^\perp \leq n - k$. □

Satz 16. *Es sei b eine symmetrische Bilinearform auf einem endlichdimensionalen Vektorraum V und U, W Unterräume von V.*

(i) Es gilt $(U + W)^\perp = U^\perp \cap W^\perp$.

(ii) Ist b nicht ausgeartet, so gilt $W^{\perp\perp} = W$.

(iii) Ist b nicht ausgeartet, so gilt $(U \cap W)^\perp = U^\perp + W^\perp$.

Beweis. (i) Ist ein Vektor $\mathbf{x}$ orthogonal zu allen $\mathbf{u} \in U$ und allen $\mathbf{w} \in W$, so ist er auch orthogonal zu allen $\mathbf{u} + \mathbf{w}$. Die Umkehrung folgt, indem wir $\mathbf{u}$ oder $\mathbf{w}$ gleich $\mathbf{0}$ setzen.

(ii) Für alle $\mathbf{w} \in W$ und $\mathbf{x} \in W^\perp$ gilt wegen der Symmetrie von b, dass $b(\mathbf{x}, \mathbf{w}) = 0$, also ist $\mathbf{w} \in W^{\perp\perp}$ und $W \subseteq W^{\perp\perp}$. Nach Lemma 6(ii) ist

$$\dim W^{\perp\perp} = \dim V - \dim W^\perp = \dim V - (\dim V - \dim W) = \dim W,$$

und mit Folgerung 5 erhalten wir die Behauptung.

(iii) Wenden wir (i) auf $U^\perp$ und $W^\perp$ an, so ergibt sich mit (ii)

$$(U^\perp + W^\perp)^\perp = U \cap W.$$

Bilden wir auf beiden Seiten das Orthogonalkomplement, so folgt wieder mit (ii) die Behauptung. □

Folgerung 7. *Ist W ein Unterraum mit der Eigenschaft $W \cap W^\perp = \{\mathbf{0}\}$, so ist V die direkte Summe von W und $W^\perp$. Die Bedingungen sind automatisch erfüllt, wenn b definit ist.*

Der Unterraum $W + W^\perp$ ist dann nämlich nach Lemma 4 die direkte Summe von W und $W^\perp$, nach Lemma 6 und Folgerung 6 ist $\dim W + W^\perp \geq \dim V$, und mit Folgerung 5 ergibt sich die Behauptung.

Zur praktischen Bestimmung von Orthogonalkomplementen wählt man meist eine Basis von V und betrachtet die Gramsche Matrix $\mathbf{B}$ von b. Nehmen wir der Einfachheit halber an, dass $V = K^n$ ist. Wie wir gesehen haben, gilt dann $b(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{B}\mathbf{y}$. Wird nun ein Unterraum W von Vektoren $\mathbf{v}_1, \ldots, \mathbf{v}_k$ erzeugt, so ist genau dann $\mathbf{x} \in W^\perp$, wenn

$$\begin{aligned} \mathbf{v}_1^\perp \mathbf{B}\mathbf{x} &= 0, \\ &\vdots \\ \mathbf{v}_k^\perp \mathbf{B}\mathbf{x} &= 0. \end{aligned}$$

Dies ist ein homogenes lineares Gleichungssystem. Fügt man die Spalten $\mathbf{v}_1$, $\ldots$, $\mathbf{v}_n$ zu einer Matrix $\mathbf{A}$ zusammen, so kann man es auch in der Form

$$\mathbf{A}^\perp \mathbf{B}\mathbf{x} = 0$$

schreiben.

Eine geometrische Interpretation der Orthogonalität bezüglich einer nicht ausgearteten Bilinearform b wird in Aufgabe 24 beschrieben. Dazu betrachtet man die algebraische Hyperfläche

$$X = \{\mathbf{x} \in V \mid q(\mathbf{x}) = 0\}.$$

Eine Gerade (einen eindimensionalen affinen Unterraum) durch einen Punkt $\mathbf{x}$ mit dem Richtungsvektor $\mathbf{y} \neq \mathbf{0}$ kann man in der Form

$$g = \{\mathbf{x} + t\mathbf{y} \mid t \in K\}$$

parametrisieren. Setzen wir diese Parametrisierung in q ein, so erhalten wir ein quadratisches Polynom in t, also hat g im Allgemeinen zwei Schnittpunkte mit X. Gehört nun $\mathbf{x}$ selbst zu X, so ist $\mathbf{y}$ ist genau dann orthogonal zu $\mathbf{x}$, wenn g eine Tangente von X ist. Dies wird in der Aufgabe im Fall von nicht isotropem $\mathbf{y}$ bewiesen, in dem die Tangente ohne Infinitesimalrechnung charakterisiert werden kann. Die Vereinigung der Tangenten nennt man den Tangentialraum an X im Punkt $\mathbf{x}$. Das Orthogonalkomplement des eindimensionalen Unterraumes, der von $\mathbf{x}$ erzeugt wird, ist somit die Menge der Verschiebungen des Tangentialraumes im Punkt $\mathbf{x}$.

15 Die Signatur

Wenn man eine indefinite quadratische Form diagonalisiert, so müssen in der entstehenden Form sowohl positive als auch negative Koeffizienten vorkommen. Bei mehr als zwei Variablen wäre es denkbar, dass die Anzahl der positiven Koeffizienten davon abhängt, welche Substitution man benutzt. Das wird aber durch folgenden Satz von James Joseph Sylvester[8] (1814–1897) ausgeschlossen.

Satz 17 (Trägheitssatz)**.** *Diagonalisiert man eine quadratische Form mit reellen Koeffizienten durch eine umkehrbare lineare Substitution, so ist die Anzahl der positiven, der negativen und der verschwindenden Koeffizienten in der diagonalisierten Form unabhängig von der Wahl der Substitution.*

Beweis. Eine quadratische Form in n Variablen definiert eine quadratische Form auf dem Vektorraum $\mathbf{R}^n$. Wir wissen bereits, dass die Anzahl der verschwindenden Koeffizienten in einer diagonalisierten Form gleich der Dimension k des Ausartungsraumes ist, also hängt sie nicht von der Substitution ab. Wir können q durch eine lineare Substitution auf die Form

$$q = l_1^2 + \cdots + l_h^2 - l_{h+1}^2 - \ldots - l_{n-k}^2$$

bringen, wobei die $l_1, \ldots, l_{n-k}$ Linearformen auf $\mathbf{R}^n$ sind und ein Vektor, auf dem all diese Linearformen den Wert 0 annehmen, im Ausartungsraum liegt. Angenommen, es gibt eine andere Darstellung

$$q = m_1^2 + \ldots + m_i^2 - m_{i+1}^2 - \ldots - m_{n-k}^2$$

[8]Er führte übrigens den Begriff „Matrix“ in die Mathematik ein.

mit Linearformen $m_1, \ldots, m_{n-k}$. Dann gilt

$$l_1^2 + \ldots + l_h^2 + m_{i+1}^2 + \ldots + m_{n-k}^2 = m_1^2 + \ldots + m_i^2 + l_{h+1}^2 + \ldots + l_{n-k}^2.$$

Hat nun $\mathbf{x} \in \mathbf{R}^n$ die Eigenschaft

$$l_1(\mathbf{x}) = \ldots = l_h(\mathbf{x}) = m_{i+1}(\mathbf{x}) = \ldots = m_{n-k}(\mathbf{x}) = 0,$$

so ist auch $l_{h+1}(\mathbf{x}) = \ldots l_{n-k}(\mathbf{x}) = 0$, also ist $\mathbf{x}$ im Ausartungsraum. Mit der Folgerung aus Satz 14 ergibt sich

$$k \geq n - (h + (n - k - i)), \quad \text{d. h.} \quad h \geq i.$$

Durch Vertauschung der Rollen erhalten wir

$$k \geq n - (i + (n - k - h)) \quad \text{d. h.} \quad i \geq h.$$

Es folgt $h = i$. □

Damit ist die folgende Definition gerechtfertigt.

Definition 28. *Ist q eine quadratische Form, in deren Diagonalisierung i positive, j negative und k verschwindende Koeffizienten vorkommen, so nennt man das Tripel (i, j, k) die Signatur von q.*

Natürlich ist diese Definition auch sinngemäß auf eine Bilinearform b auf einem abstrakten Vektorraum und auf ihre Spezialisierung q anwendbar.

Ist V die direkte Summe von Unterräumen V_1 und V_2, die zueinander bezüglich b orthogonal sind, so hängt die Signatur (i, j, k) von q mit den Signaturen (i_1, j_1, k_1) und (i_2, j_2, k_2) ihrer Einschränkungen auf V_1 und V_2 offensichtlich wie folgt zusammen:

$$i = i_1 + i_2, \qquad j = j_1 + j_2, \qquad k = k_1 + k_2.$$

Satz 18. *Hat die quadratische Form q auf dem Vektorraum V die Signatur (i, j, k), so ist i die maximale Dimension eines Unterraums, auf dem die Einschränkung von q positiv definit ist, und j die maximale Dimension eines Unterraums, auf dem die Einschränkung von q negativ definit ist*

Beweis. Es sei W ein Unterraum, auf dem die Einschränkung von q positiv definit ist. Dann ist V nach Folgerung 7 die direkte Summe von W und $W^\perp$. Nun folgt die erste Behauptung aus der obigen Bemerkung. Genauso beweist man die zweite Behauptung. □

Folgerung 8. *Ist (i, j, k) die Signatur einer quadratischen Form auf einem reellen Vektorraum V und (i', j', k') die Signatur ihrer Einschränkung auf einen Unterraum W, so gilt $i' \leq i$ und $j' \leq j$.*

Ist b definit, so gilt das offenbar auch für die Einschränkung von b auf einen Unterraum W. Im Allgemeinen gilt aber nicht $k' \leq k$. Wir werden auf diese Frage noch zurück kommen.

16 Lineare Abbildungen

Genau so, wie wir den Begriff der (Bi-)Linearform aus der Sprache der homogenen Polynome in die Sprache der Funktionen auf Vektorräumen übersetzt haben, wollen wir nun den Begriff der linearen Substitution übersetzen.

Definition 29. *Es seien V und W Vektorräume über einem Körper K. Eine Abbildung $f : V \to W$ wird linear genannt, wenn für alle $\mathbf{x}$, $\mathbf{y} \in V$ und alle $a \in K$ gilt*

$$f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}), \qquad f(a \cdot \mathbf{x}) = a \cdot f(\mathbf{x}).$$

Im Spezialfall, wenn $W = K$ ist, erhalten wir den Begriff einer Linearform. Im allgemeinen Fall können wir eine Basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ von V und eine Basis $\mathbf{w}_1, \ldots, \mathbf{w}_m$ von W wählen und die Bilder der Basisvektoren von V durch die Basisvektoren von W ausdrücken:

$$f(\mathbf{v}_i) = a_{1i}\mathbf{w}_1 + \ldots + a_{mi}\mathbf{w}_m.$$

Hat $\mathbf{x}$ die Koordinaten $x_1, \ldots, x_n$ und $f(\mathbf{x})$ die Koordinaten $y_1, \ldots, y_n$, so ergibt sich ähnlich wie auf S. 18

$$\begin{aligned} y_1 &= a_{11}x_1 + \ldots + a_{1n}x_n \\ &\vdots \\ y_m &= a_{m1}x_1 + \ldots + a_{mn}x_n \end{aligned}$$

Dies ist in der Tat eine lineare Substitution, die durch eine Matrix $\mathbf{A}$ mit den Einträgen a_{ij} beschrieben wird. Ist $g : U \to V$ eine weitere lineare Abbildung, die bezüglich der obigen Basis von V und einer Basis $\mathbf{u}_1, \ldots, \mathbf{u}_p$ von U durch eine Matrix $\mathbf{B}$ dargestellt wird, so wird die Verkettung $f \circ g$ in den obigen Basen von W und U durch die Produktmatrix $\mathbf{AB}$ dargestellt.

Zur Beschreibung von Abbildungen eines Vektorraums V in sich selbst benutzt man gewöhnlich ein und die selbe Basis für Definitionsbereich und Zielbereich. Dann wird z. B. die identische Abbildung durch die Einheitsmatrix beschrieben. Ist in den obigen Ausführungen $U = V$, so ist $f \circ g$ genau dann die identische Abbildung, wenn $\mathbf{AB} = \mathbf{E}$ ist. Eine lineare Abbildung ist also genau dann umkehrbar, wenn Definitions- und Zielbereich die selbe Dimension haben und die beschreibende Matrix regulär ist.

Beispiel. Ist V die direkte Summe von Unterräumen U und W, so definieren wir die Projektion p von V auf W längs U wie folgt. Jeder Vektor $\mathbf{x} \in V$ lässt sich auf eindeutige Weise in der Form $\mathbf{x} = \mathbf{u} + \mathbf{w}$ schreiben, wobei $\mathbf{u} \in U$ und $\mathbf{w} \in W$, und für dieses $\mathbf{x}$ setzen wir dann $p(\mathbf{x}) = \mathbf{w}$.

Nun bringen wir symmetrische Bilinearformen ins Spiel. Aus der elementaren Geometrie ist folgendes Beispiel bekannt.
Beispiel. Auf dem reellen Vektorraum V sei ein Skalarprodukt, d. h. eine symmetrische Bilinearform b mit Spezialisierung q gegeben. Ist W ein Unterraum, so ist V die direkte Summe von W und dem Orthogonalkomplement $W^\perp$. Die Projektion p längs $W^\perp$ heißt dann Orthogonalprojektion von V auf W. Für ein beliebiges $\mathbf{x} \in V$ ist dann $p(\mathbf{x})$ unter allen Vektoren $\mathbf{w} \in W$ derjenige, für den $\|\mathbf{x} - \mathbf{w}\|$ den kleinsten Wert annimmt. Wegen $\mathbf{x} - p(\mathbf{x}) \in W^\perp$ ist nämlich

$$q(\mathbf{x}-\mathbf{w}) = q((\mathbf{x}-p(\mathbf{x}))+(p(\mathbf{x})-\mathbf{w})) = q(\mathbf{x}-p(\mathbf{x}))+q(p(\mathbf{x})-\mathbf{w}) \leq q(p(\mathbf{x})-\mathbf{x}),$$

wobei Gleichheit genau dann eintritt, wenn $q(p(\mathbf{x}) - \mathbf{w}) = 0$.

Ist V ein Vektorraum über einem beliebigen Körper K, so gibt es keinen Begriff der Definitheit, und für manche Unterräume ergibt die Orthogonalprojektion keinen Sinn.

Definition 30. *Ein Unterraum W von V heißt isotrop bezüglich der symmetrischen Bilinearform b, wenn die Einschränkung von b auf W ausgeartet ist.*

Offensichtlich ist W genau dann nicht isotrop, wenn $W \cap W^\perp = \{\mathbf{0}\}$. Unter dieser Bedingung ist V nach Folgerung 7 die direkte Summe von W und $W^\perp$, und die Orthogonalprojektion auf W ist definiert. Ist W nicht isotrop und b nicht ausgeartet, so ist mit Satz 16(ii) auch $W^\perp$ nicht isotrop.

Wird W von einem einzigen nicht isotropen Vektor $\mathbf{v}$ erzeugt, so ist $q(\mathbf{x}) = a\mathbf{v}$ mit einer reellen Zahl a. Die Bedingung $\mathbf{x} - p(\mathbf{x}) \in W^\perp$ bedeutet nun

$$b(\mathbf{x} - a\mathbf{v}, \mathbf{v}) = 0, \qquad \text{also} \qquad b(\mathbf{x}, \mathbf{v}) = aq(\mathbf{v}).$$

Daraus ergibt sich die explizite Formel

$$p(\mathbf{x}) = \frac{b(\mathbf{x}, \mathbf{v})}{q(\mathbf{v})}\mathbf{v}$$

für die Orthogonalprojektion auf den Vektor $\mathbf{v}$.

In der Geometrie sind vor allem lineare Abbildungen wichtig, bei denen die Norm erhalten bleibt. An Stelle von Skalarprodukten betrachten wir beliebige symmetrische Bilinearformen.

Definition 31. *Es seien V und W Vektorräume über einem Körper K, b eine symmetrische Bilinearform auf V und c eine ebensolche auf W. Eine*

lineare Abbildung $f : V \to W$ heißt Isometrie bezüglich b und c, wenn für alle $\mathbf{x}, \mathbf{y} \in V$ gilt

$$c(f(\mathbf{x}), f(\mathbf{y})) = b(\mathbf{x}, \mathbf{y}).$$

Gibt es eine umkehrbare Isometrie bezüglich b und c, so heißen b und c äquivalent.

Beispiel. Es sei W ein nicht isotroper Unterraum von V. Die Spiegelung von V in W ist die folgendermaßen definierte Abbildung $s : V \to V$. Da V die direkte Summe von W und $W^\perp$ ist, lässt sich jeder Vektor $\mathbf{x} \in V$ in der Form $\mathbf{x} = \mathbf{w} + \mathbf{u}$ schreiben, wobei $\mathbf{w} \in W$ und $\mathbf{u} \in W^\perp$. Wir setzen $s(\mathbf{x}) = \mathbf{w} - \mathbf{u}$. Wegen $b(\mathbf{w}, \mathbf{u}) = 0$ gilt

$$q(\mathbf{w} + \mathbf{u}) = q(\mathbf{w}) + q(\mathbf{u}) = q(\mathbf{w} - \mathbf{u}),$$

also ist s eine Isometrie bezüglich b, und wegen $s(s(\mathbf{x})) = \mathbf{x}$ ist sie umkehrbar.

Angenommen, $W^\perp$ wird von einem einzigen nicht isotropen Vektor $\mathbf{v}$ erzeugt. Dann ist $\mathbf{u}$ die Projektion von $\mathbf{x}$ auf $\mathbf{v}$. Setzen wir die obige Formel in $s(\mathbf{x}) = \mathbf{x} - 2\mathbf{u}$ ein, so erhalten wir

$$s(\mathbf{x}) = \mathbf{x} - 2\frac{b(\mathbf{x}, \mathbf{v})}{q(\mathbf{v})}\mathbf{v}.$$

Dies definiert eine umkehrbare Isometrie für jeden nicht isotropen Vektor $\mathbf{v}$.

Sind Basen der Vektorräume V und W gegeben, so wird eine lineare Abbildung $f : V \to W$ durch eine Matrix $\mathbf{A}$ gegeben, und die Bilinearformen b und c haben bezüglich dieser Basen Gramsche Matrizen $\mathbf{B}$ und $\mathbf{C}$. Wir nehmen einmal der Einfachheit halber an, dass V und W selbst Räume von Spaltenvektoren sind. Dann ist

$$f(\mathbf{x}) = \mathbf{A}\mathbf{x}, \qquad b(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{B}\mathbf{y}, \qquad c(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\top \mathbf{C}\mathbf{v},$$

und es folgt

$$c(f(\mathbf{x}), f(\mathbf{y})) = (\mathbf{A}\mathbf{x})^\top \mathbf{C}(\mathbf{A}\mathbf{y}).$$

Die Abbildung f ist also genau dann eine Isometrie bezüglich b und c, wenn

$$\mathbf{A}^\top \mathbf{C}\mathbf{A} = \mathbf{B}.$$

Letzteres gilt auch ohne die vereinfachende Annahme.

Da man auch einen Basiswechsel als lineare Abbildung interpretieren kann, sollte die Ähnlichkeit zu den Ausführungen auf S. 48 nicht verwundern.

Eng verwandt mit linearen Abbildungen sind affine Abbildungen:

Definition 32. *Es seien A und B affine Räume. Eine Abbildung $g : A \to B$ heißt affine Abbildung, wenn sie Parallelogramme auf Parallelogramme abbildet. Mit anderen Worten, für beliebige Punkte P, Q, R und S von A mit der Eigenschaft $P \begin{matrix} Q \\ R \end{matrix} S$ muss gelten $g(P) \begin{matrix} g(Q) \\ g(R) \end{matrix} g(S)$.*

Sind V und W die Vektorräume der Verschiebungen von A bzw. B, so erhalten wir eine Abbildung $f : V \to W$, indem wir setzen

$$f\big(\overrightarrow{PQ}\big) = \overrightarrow{g(P)g(Q)}.$$

Sie hat die Eigenschaft

$$f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$$

für alle $\mathbf{x}$, $\mathbf{y} \in V$.

Definition 33. *Es seien A und B reelle affine Räume. Eine affine Abbildung $g : A \to B$ ist eine reelle affine Abbildung, wenn die zugehörige Abbildung $f : V \to W$ zwischen den Vektorräumen ihrer Verschiebungen eine lineare Abbildung ist.*

Man kann jeden reellen Vektorraum als affinen Raum betrachten, und dann sind alle linearen Abbildungen auch reelle affine Abbildungen. Es gibt aber reelle affine Abbildungen zwischen Vektorräumen, die nicht linear sind, wie z. B. die für einen festen Vektor $\mathbf{v} \in V$ durch $g(\mathbf{x}) = \mathbf{x} + \mathbf{v}$ gegebene Abbildung $g : V \to V$, die man eine Verschiebung nennt.

Nun können wir endlich den Begriff der Bewegung streng definieren, der im Schulunterricht eine wichtige Rolle spielt.

Definition 34. *Es sei A ein Euklidischer Raum mit der Abstandsfunktion d und B ein Euklidischer Raum mit der Abstandsfunktion e. Eine Abbildung $g : A \to B$ heißt Isometrie, wenn für alle Punkte P und Q von A gilt*

$$e\big(g(P), g(Q)\big) = d(P, Q).$$

Eine umkehrbare Isometrie $A \to A$ heißt Bewegung von A.

Mit Hilfe von Aufgabe 50* kann man zeigen, dass jede Isometrie zwischen Euklidischen Räumen eine reelle affine Abbildung ist.

17 Der Satz von Witt

Im Folgenden sei b eine symmetrische Bilinearform auf einem Vektorraum V über einem Körper K, in dem 2 invertierbar ist, und q die Spezialisierung von b. Ist $b \neq 0$, so gibt es immer einen von Null verschiedenen nichtisotropen Vektor, denn sonst wäre die Polarisierung $2b$ von q gleich Null. Der Kürze halber führen wir noch folgende Bezeichnung ein.

Definition 35. *Es seien U_1 und U_2 Unterräume von V. Unter einer Isometrie $U_1 \to U_2$ bezüglich b verstehen wir eine Isometrie bezüglich der Einschränkungen von b auf U_1 und auf U_2. Die Unterräume U_1 und U_2 werden bezüglich b äquivalent genannt, wenn es eine umkehrbare Isometrie $U_1 \to U_2$ bezüglich b gibt, d. h. wenn die beiden Einschränkungen von b äquivalent sind.*

Folgender Satz wurde von Ernst Witt (1911–1991) bewiesen.

Satz 19. *Sind U_1 und U_2 nichtisotrope und bezüglich b äquivalente Unterräume von V, so sind auch $U_1^\perp$ und $U_2^\perp$ bezüglich b äquivalent.*

Beweis. Laut Definition gibt es eine umkehrbare Isometrie $f : U_1 \to U_2$ bezüglich b. Wir beweisen den Satz durch Induktion nach der natürlichen Zahl $k = \dim U_1 = \dim U_2$. Für $k = 0$ ist nichts zu beweisen. Wir betrachten nun den Fall, dass $k = 1$ ist. Dann wählen wir einen von Null verschiedenen Vektor $\mathbf{u}_1 \in U_1$ und setzen $\mathbf{u}_2 = f(\mathbf{u}_1)$, so dass $q(\mathbf{u}_1) = q(\mathbf{u}_2)$. Nach der Parallelogramm-Identität (Präsenzaufgabe 22) gilt

$$q(\mathbf{u}_1 + \mathbf{u}_2) + q(\mathbf{u}_1 - \mathbf{u}_2) = 2q(\mathbf{u}_1) + 2q(\mathbf{u}_2) = 4q(\mathbf{u}_1),$$

was nicht Null ist, weil U_1 nicht isotrop ist. Somit können $q(\mathbf{u}_1 + \mathbf{u}_2)$ und $q(\mathbf{u}_1 - \mathbf{u}_2)$ nicht beide gleich Null sein. Indem wir notfalls $\mathbf{u}_2$ durch $-\mathbf{u}_2$ ersetzen, können wir annehmen, dass der Vektor $\mathbf{v} = \mathbf{u}_1 - \mathbf{u}_2$ nicht isotrop ist. Dann wird durch

$$s(\mathbf{x}) = \mathbf{x} - 2\frac{b(\mathbf{x}, \mathbf{v})}{q(\mathbf{v})}\mathbf{v}$$

eine Spiegelung $s : V \to V$ definiert. Es gilt

$$q(\mathbf{v}) = q(\mathbf{u}_1) + q(\mathbf{u}_2) - 2b(\mathbf{u}_1, \mathbf{u}_2) = 2(q(\mathbf{u}_1) - b(\mathbf{u}_1, \mathbf{u}_2)) = 2b(\mathbf{u}_1, \mathbf{v}),$$

also

$$s(\mathbf{u}_1) = \mathbf{u}_1 - \mathbf{v} = \mathbf{u}_2.$$

Die umkehrbare Isometrie s von V auf sich selbst bezüglich b bildet also U_1 auf U_2 und somit $U_1^\perp$ auf $U_2^\perp$ ab. Ihre Einschränkung ist eine umkehrbare Isometrie $U_1^\perp \to U_2^\perp$ bezüglich b, und der Induktionsanfang ist abgeschlossen.

Nun kommen wir zum Induktionsschritt. Es sei also $k > 1$, und der Satz gelte bereits für Unterräume kleinerer Dimension. Da die Einschränkung von b auf den Unterraum U_1 nicht Null ist, enthält dieser einen nicht isotropen Vektor $\mathbf{u}_1 \neq \mathbf{0}$. Der von $\mathbf{u}_1$ erzeugte Unterraum V_1 ist ebenso wie sein Bild $V_2 = f(V_1)$ nicht isotrop. Laut Induktionsanfang gibt es eine umkehrbare Isometrie $h : V_1^\perp \to V_2^\perp$ bezüglich b.

Das Orthogonalkomplement von V_1 bezüglich der nicht ausgearteten Einschränkung von b auf U_1 ist der nicht isotrope Unterraum $W_1 = V_1^\perp \cap U_1$. Da

U_1 die direkte Summe von V_1 und W_1 ist, gilt $\dim W_1 = k - 1$. Ähnlich ist $W_2 = f(W_1)$ das Orthogonalkomplement von V_2 bezüglich der Einschränkung von b auf U_2, also $W_2 = V_2^\perp \cap U_2$.

Das Orthogonalkomplement von W_1 bezüglich der Einschränkung von b auf $V_1^\perp$ ist

$$W_1^\perp \cap V_1^\perp = (W_1 + V_1)^\perp = U_1^\perp,$$

wobei wir Lemma 16 benutzt haben. Analog ist das Orthogonalkomplement von W_2 bezüglich der Einschränkung von b auf $V_2^\perp$ gleich $U_2^\perp$, während das von $h(W_1)$ gleich $h(U_1^\perp)$ ist, denn h ist eine Isometrie.

Die Unterräume $W_2 = f(W_1)$ und $h(W_1)$ sind äquivalent bezüglich b, also auch bezüglich der Einschränkung von b auf $V_2^\perp$. Nach Induktionsvoraussetzung gilt das auch für ihre Orthogonalkomplemente $U_2^\perp$ und $h(U_1^\perp)$, die somit auch bezüglich b äquivalent sind. Der Unterraum $h(U_1^\perp)$ ist bezüglich b äquivalent zu $U_1^\perp$, und die Behauptung folgt mit der Transitivität der Äquivalenz. □

Für isotrope Unterräume gilt der Satz leider nicht.

Definition 36. *Ein Unterraum W von V heißt vollständig isotrop, wenn die Einschränkung von b auf W gleich Null ist. Wir sagen, dass b zerfällt, wenn b nicht ausgeartet ist und V die direkte Summe von zwei vollständig isotropen Unterräumen ist.*

Offensichtlich ist W genau dann vollständig isotrop, wenn $W \subseteq W^\perp$. Ist V die direkte Summe von zueinander orthogonalen Unterräumen V_1 und V_2 und zerfallen die Einschränkungen von b auf V_1 und V_2, so zerfällt auch b. Ein Beispiel für eine zerfallende Form auf K^2 haben wir bereits kennengelernt, nämlich $b(\mathbf{x}, \mathbf{y}) = x_1y_2 + x_2y_1$. Man sieht leicht, dass jede zerfallende Bilinearform auf einem zweidimensionalen Vektorraum zu dieser Bilinearform äquivalent ist. Mehr noch:

Lemma 7. *Es sei b nicht ausgeartet und U ein vollständig isotroper Unterraum.*

(i) Es gibt einen vollständig isotropen Unterraum W, so dass $U \cap W = \{\mathbf{0}\}$, $\dim U = \dim W$ und die Einschränkung von b auf $U + W$ nicht ausgeartet ist.

(ii) Ist außerdem $\dim U = \frac{1}{2} \dim V$, so ist V eine orthogonale direkte Summe von zweidimensionalen Unterräumen, auf denen die Einschränkung von b zerfällt.

Beweis. Wir beweisen zunächst (i) durch vollständige Induktion nach $k = \dim U$. Ist $k = 0$, so können wir $W = \{\mathbf{0}\}$ setzen. Nun sei $k > 0$. Dann gibt es einen Vektor $\mathbf{u} \neq \mathbf{0}$ in U. Da b nicht ausgeartet ist, gibt es einen Vektor $\mathbf{v}$, so dass $b(\mathbf{u}, \mathbf{v}) \neq 0$. Wir können annehmen, dass $b(\mathbf{u}, \mathbf{v}) = 1$. Setzen wir $\mathbf{w} = \mathbf{v} + a\mathbf{u}$ mit $a \in K$, so ist $b(\mathbf{u}, \mathbf{w}) = 1$ und

$$q(\mathbf{w}) = 2a + q(\mathbf{v}).$$

Da 2 invertierbar ist, können wir a so wählen, dass $q(\mathbf{w}) = 0$. Die linear unabhängigen Vektoren $\mathbf{u}$ bzw. $\mathbf{w}$ erzeugen eindimensionale Unterräume U' und W', und die Einschränkung von b auf den Unterraum $V' = U' + W'$ zerfällt.

Nun sei $V'' = V'^{\perp}$. Wegen $\mathbf{u} \notin V''$ hat $U'' = U \cap V''$ kleinere Dimension als U. Außerdem gilt $U = U' + U''$, denn für $\mathbf{x} \in U$ ist $\mathbf{x} - b(\mathbf{x}, \mathbf{w})\mathbf{u}$ orthogonal zu $\mathbf{u}$ und $\mathbf{w}$, also in V''. Laut Induktionsvoraussetzung gibt es einen vollständig isotropen Unterraum W'' von V'', so dass $U'' \cap W'' = \{\mathbf{0}\}$ und die Einschränkung von b auf $U'' + W''$ nicht ausgeartet ist. Setzen wir $W = W' + W''$, so folgt die Behauptung (i).

Nun beweisen wir (ii) durch vollständige Induktion nach k. Wieder ist für $k = 0$ nichts zu beweisen. Für $k > 0$ liefert der Beweis von Teil (i) eine Zerlegung von V in eine orthogonale direkte Summe eines zweidimensionalen Unterraums V', auf dem die Einschränkung von b zerfällt, und eines Unterraumes V'', der einen vollständig isotropen Unterraum U'' enthält, wobei $\dim U'' = k - 1 = \frac{1}{2} \dim V''$. Nach Induktionsvoraussetzung ist V'' eine orthogonale direkte Summe von zweidimensionalen Unterräumen, auf denen die Einschränkung von b zerfällt. □

Folgerung 9. *Zwei zerfallende quadratische Formen b_1 und b_2 auf Vektorräumen V_1 und V_2 sind genau dann äquivalent, wenn $\dim V_1 = \dim V_2$. Ist V_i die direkte Summe von vollständig isotropen Unterräumen U_i und W_i, so gibt es eine Isometrie $f : V_1 \to V_2$ bezüglich b_1 und b_2, so dass $f(U_1) = U_2$ und $f(W_1) = W_2$.*

Definition 37. *Eine symmetrische Bilinearform heißt anisotrop, wenn der einzige isotrope Vektor der Nullvektor ist.*

Folgerung 10. *Ist b eine symmetrische Bilinearform auf einem Vektorraum V, dann gibt es Unterräume V_1 und V_2, so dass V die orthogonale direkte Summe von V_1, V_2 und $V^{\perp}$ ist, wobei die Einschränkung von b auf V_1 zerfällt und die Einschränkung auf V_2 anisotrop ist.*

Beweis. Nach Satz 15 gibt es einen Unterraum V', so dass V die direkte Summe von V' und $V^{\perp}$ ist. Wegen $V' \cap V'^{\perp} \subset V' \cap V^{\perp} = \{\mathbf{0}\}$ ist die Einschränkung

von b auf V' nicht ausgeartet. Nach Lemma 5 muss jede aufsteigende Folge von vollständig isotropen Unterräumen $U_1 \subset U_2 \subset \ldots$ in V' irgendwann abbrechen, da V' endliche Dimension hat. Es gibt also in V' einen maximalen[9] vollständig isotropen Unterraum U, und nach Lemma 7 finden wir einen Unterraum W, so dass die Einschränkung von b auf $V_1 = U + W$ zerfällt. Nach Folgerung 7 ist V' die direkte Summe von V_1 und $V_2 = V' \cap V_1^\perp$. Ist $\mathbf{u} \in V_2$ ein isotroper Vektor, so ist der von U und $\mathbf{u}$ erzeugte Unterraum vollständig isotrop, und wegen der Maximalität von U folgt $\mathbf{u} = 0$. Somit ist die Einschränkung von b auf V_2 anisotrop. □

Die Unterräume V_1 und V_2 sind im Unterschied zu $V^\perp$ nicht eindeutig bestimmt. Wie steht es mit ihren Dimensionen?

Satz 20. *Ist b nicht ausgeartet und sind U_1 und U_2 vollständig isotrope Unterräume gleicher Dimension, so gibt es eine umkehrbare Isometrie $g : V \to V$ bezüglich b, die U_1 auf U_2 abbildet.*

Beweis. Wir finden Unterräume W_1 und W_2 wie im Lemma und setzen $V_i = U_i + W_i$. Nach der Folgerung 9 gibt es eine umkehrbare Isometrie $f : V_1 \to V_2$ bezüglich b, so dass $f(U_1) = U_2$. Nach Satz 19 gibt es eine umkehrbare Isometrie $h : V_1^\perp \to V_2^\perp$ bezüglich b. Da V die direkte Summe von V_i und $V_i^\perp$ ist, können wir die gesuchte Isometrie g durch

$$g(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + h(\mathbf{y})$$

für $\mathbf{x} \in V_1$ und $\mathbf{y} \in V_1^\perp$ definieren. □

Folgerung 11. *Alle maximalen vollständig isotropen Unterräume haben die gleiche Dimension.*

Beweis. Es seien U_1 und U_2 maximale vollständig isotrope Unterräume. Da beide den Ausartungsraum enthalten, genügt es in den Bezeichnungen von Folgerung 10 zu zeigen, dass $\dim U_1 \cap V' = \dim U_2 \cap V'$. Wir können also ohne Beschränkung der Allgemeinheit annehmen, dass b nicht ausgeartet ist. Ist nun z. B. $\dim U_1 \geq \dim U_2$, so gibt es einen Unterraum U_1' von U_1, so dass $\dim U_1' = \dim U_2$. Natürlich ist auch U_1' vollständig isotrop, und nach Satz 20 gibt es eine umkehrbare Isometrie g, so dass $g(U_1') = U_2$, also $U_2 \subseteq g(U_1)$. Da g eine Isometrie ist, ist auch $g(U_1)$ vollständig isotrop, und aus der Maximalität von U_2 folgt $g(U_1) = U_2$. Wegen der Umkehrbarkeit von g ist $\dim U_1 = \dim U_2$. □

[9] der also in keinem vollständig isotropen Unterraum echt enthalten ist

Definition 38. *Die maximale Dimension eines vollständig isotropen Unterraums bezüglich einer symmetrischen Bilinearform heißt Index (manchmal auch Witt-Index) dieser Bilinearform.*

Da jeder vollständig isotrope Unterraum eines Unterraums auch ein vollständig isotroper Unterraum des gesamten Vektorraums ist, erhalten wir:

Folgerung 12. *Der Index der Einschränkung einer symmetrischen Bilinearform b auf einen Unterraum ist nicht größer als der Index von b.*

Ist V die orthogonale direkte Summe von Unterräumen, so ergibt sich der Index von b im Allgemeinen nicht als Summe der Indizes der Einschränkungen von b, wie das Beispiel $q(x_1, x_2) = x_1^2 - x_2^2$ zeigt.

Folgerung 13. *In den Bezeichnungen von Folgerung 10 ist der Index von b gleich*

$$\frac{1}{2}\dim V_1 + \dim V^\perp.$$

Die Einschränkung von b auf V_2 ist bis auf Äquivalenz eindeutig bestimmt; man nennt sie den anisotropen Kern von b.

Die erste Behauptung folgt aus Satz 20 und Lemma 7. Nach Folgerung 9 ist die Einschränkung von b auf V_1, also auch auf $V_1 + V^\perp$, bis auf Äquivalenz eindeutig bestimmt, und die zweite Behauptung folgt aus Satz 19.

Wir betrachten nun den Fall eines reellen Vektorraums V. Bezüglich einer geeigneten Orthogonalbasis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ erhalten wir die quadratische Form

$$x_1^2 + \ldots + x_i^2 - x_{i+1}^2 - \ldots - x_{i+j}^2.$$

Die Vektoren $\mathbf{v}_{i+j+1}, \ldots, \mathbf{v}_n$ erzeugen den Ausartungsraum. Die Vektoren $\mathbf{v}_l$ und $\mathbf{v}_{i+l}$ für alle Indizes l mit den Eigenschaften $l \leq i$ und $l \leq j$ erzeugen einen Unterraum, auf dem die Einschränkung von b zerfällt. Die verbleibenden Basisvektoren erzeugen einen Unterraum, auf dem die Einschränkung von b definit und somit anisotrop ist. Wir erhalten:

Folgerung 14. *Es sei b eine symmetrische Bilinearform mit der Signatur (i, j, k) auf einem reellen Vektorraum. Dann ist der Index von b gleich*

$$\min(i, j) + k.$$

Mit Folgerung 12 ergibt sich eine zusätzliche Bedingung zu den bereits in Folgerung 8 formulierten Bedingungen, denen die Signatur der Einschränkung einer symmetrischen Bilinearform genügen muss. Hier sind z. B. für jede Signatur auf einem dreidimensionalen Raum V die möglichen Signaturen der Einschränkung auf einen zweidimensionalen Unterraum W:

V	W
$(3,0,0)$	$(2,0,0)$
$(2,1,0)$	$(2,0,0)$, $(1,1,0)$, $(1,0,1)$
$(2,0,1)$	$(2,0,0)$, $(1,0,1)$
$(1,1,1)$	$(1,1,0)$, $(1,0,1)$, $(0,1,1)$, $(0,0,2)$
$(1,0,2)$	$(1,0,1)$, $(0,0,2)$
$(0,0,3)$	$(0,0,0)$

(Man erhält die fehlenden Signaturen, indem man das Vorzeichen der Form ändert, also i und j vertauscht.) Für reelle Vektorräume beliebiger Dimension sind die genannten Bedingungen noch nicht hinreichend; die endgültige Antwort findet man in Aufgabe 55.

Mit Hilfe dieser Überlegungen kann man Informationen über die Durchschnitte zwischen einer algebraischen Fläche X in einem reellen affinen Raum A und Ebenen B in A gewinnen. Wir erinnern uns, dass X als Nullstellenmenge einer polynomialen Funktion p definiert war, wobei wir nur Polynome vom Grad 2 betrachtet haben. Wählen wir einen Koordinatenursprung $O \in A$, so können wir p als Funktion von Ortsvektoren betrachten und in homogene Komponenten zerlegen. Die Komponente q vom Grad 2 ist eine quadratische Form, deren Signatur bereits eine grobe Klassifikation von X ermöglicht. Der Schnitt von X mit der Ebene B ist die Nullstellenmenge der Einschränkung von p, und die Signatur der Einschränkung von q liefert bereits wichtige Informationen.

Weitere Betrachtungen, die wir hier nicht ausführen wollen, liefern folgende Möglichkeiten:

Fläche X	ebene Schnitte $X \cap B$
Ellipsoid	Ellipsen, Punkte, leere Mengen
einschaliges Hyperboloid	alles außer Punkten und leeren Mengen
Doppelkegel	alles außer zwei parallelen Geraden und leeren Mengen
zweischaliges Hyperboloid	alles außer zwei Geraden
elliptisches Paraboloid	Parabeln, Ellipsen, Punkte, leere Mengen
hyperbolisches Paraboloid	Parabeln, Hyperbeln, zwei sich schneidende Geraden

Da sich fast alle Kurven zweiter Ordnung als ebene Schnitte eines Doppelkegels realisieren lassen, der sich leicht geometrisch konstruieren lässt, werden sie seit dem griechischen Altertum Kegelschnitte genannt.

18 Reduktion

In diesem Abschnitt interessieren uns quadratische Formen mit ganzzahligen Koeffizienten. Da dies ein schwieriges Thema ist, beschränken wir uns auf quadratische Formen in zwei Variablen, die wir in der Form

$$q(x,y) = ax^2 + bxy + cy^2 \tag{10}$$

mit ganzen Zahlen a, b und c schreiben können. Klassische Ergebnisse der Zahlentheorie geben an, welche ganzen Zahlen als Werte einer solchen Form angenommen werden. Wir werden uns statt dessen mit der Frage der Klassifikation befassen, die näher an unserem bisherigen Stoff liegt. Die Polarisierung von q ist übrigens

$$p(r,s,t,u) = q(r+t,s+u) - q(r,s) - q(t,u) = 2art + b(ru+st) + 2csu.$$

Die ganze Zahl $d = b^2 - 4ac$ bezeichnet man als Diskriminante[10] der quadratischen Form q. Wir behaupten, dass Formen mit negativer Diskriminante definit und Formen mit positiver Diskriminante indefinit sind, während Formen mit verschwindender Diskriminante ausgeartet sind. Dies beweist man im Fall $a \neq 0$ mit der Methode der quadratischen Ergänzung:

$$q(x,y) = a\left(x + \frac{b}{2a}y\right)^2 + \left(c - \frac{b^2}{4a}\right)y^2 = a\left(\left(x + \frac{b}{2a}y\right)^2 - d\left(\frac{y}{2a}\right)^2\right),$$

Im Fall $a = 0$, aber $c \neq 0$ vertauscht man einfach die Rollen von x und y, und im Fall $a = c = 0$ ist die Behauptung offensichtlich.

Wir können auch quadratische Formen q_0 im Sinne von Definition 11 als Funktionen auf einem freien Modul V über dem Ring $\mathbf{Z}$ der ganzen Zahlen betrachten. Ist V ein freier Modul vom Rang 2 und wählen wir eine Basis $\mathbf{v}$, $\mathbf{w}$, so ist

$$q_0(x\mathbf{v} + y\mathbf{w}) = q(x,y),$$

wobei q eine quadratische Form im Sinne von Polynomen ist.

Wir erinnern uns daran, wie sich diese Darstellung beim Übergang zu einer anderen Basis $\mathbf{v}'$, $\mathbf{w}'$ ändert. Es gibt ganze Zahlen r, s, t und u, so dass

$$\mathbf{v}' = r\mathbf{v} + s\mathbf{w}$$
$$\mathbf{w}' = t\mathbf{v} + u\mathbf{w}$$

[10] Die Zahl $\frac{d}{(2a)^2}$ ist die Diskriminante des Polynoms $\frac{q(x,1)}{a}$, und die Zahl $\frac{d}{(2c)^2}$ ist die Diskriminante des Polynoms $\frac{q(1,y)}{c}$.

Ist $x\mathbf{v} + y\mathbf{w} = x'\mathbf{v}' + y'\mathbf{w}'$, so egibt sich die umkehrbare lineare Substitution

$$\begin{aligned} x &= rx' + ty' \\ y &= sx' + uy' \end{aligned}$$

und durch Einsetzen erhalten wir $q(x, y) = q'(x', y')$, wobei

$$q'(x', y') = a'x'^2 + b'x'y' + c'y'^2$$

mit

$$a' = q(r, s), \qquad b' = p(r, s, t, u), \qquad c' = q(t, u).$$

Die Diskriminante von q' ergibt sich nach längerer Rechnung zu

$$d' = b'^2 - 4a'c' = (ru - st)^2 d.$$

Die ganze Zahl $ru - st$ nennt man übrigens die Determinante der Matrix

$$\begin{pmatrix} r & t \\ s & u \end{pmatrix}.$$

Man kann den Basiswechsel der Koordinaten auch durch

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} r & t \\ s & u \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix}$$

beschreiben.

Gehen wir von $\mathbf{v}'$, $\mathbf{w}'$ zu einer weiteren Basis $\mathbf{v}''$, $\mathbf{w}''$ über, so gilt für die Koordinaten

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} r' & t' \\ s' & u' \end{pmatrix} \begin{pmatrix} x'' \\ y'' \end{pmatrix},$$

mit ganzen Zahlen r', s', t' und u', und der durchgehende Basiswechsel von $\mathbf{v}$, $\mathbf{w}$ zu $\mathbf{v}''$, $\mathbf{w}''$ wird durch die Matrix

$$\begin{pmatrix} r'' & t'' \\ s'' & u'' \end{pmatrix} = \begin{pmatrix} r & t \\ s & u \end{pmatrix} \begin{pmatrix} r' & t' \\ s' & u' \end{pmatrix}$$

beschrieben. Man rechnet leicht nach, dass

$$r''u'' - s''t'' = (ru - st)(r'u' - s't'),$$

mit anderen Worten, die Determinante eines Produktes von Matrizen ist das Produkt der Determinanten.

Das Gesagte gilt insbesondere, wenn $\mathbf{v}'' = \mathbf{v}$ und $\mathbf{w}'' = \mathbf{w}$ ist. In diesem Fall ist die Produktmatrix die Einheitsmatrix, deren Determinante gleich 1

ist. Wir sehen, dass das Produkt der Determinanten gleich 1 ist. Da diese Determinanten ganze Zahlen sind, folgt

$$ru - st = \pm 1$$

und

$$d' = d.$$

Quadratische Formen q und q', die wie hier durch eine umkehrbare lineare Substitution zusammenhängen, nennt man äquivalent, weil sie äquivalente quadratische Formen auf dem Modul $\mathbf{Z}^2$ definieren. Wir haben gesehen, dass äquivalente Formen die gleiche Diskriminante besitzen.

Wir wollen nun versuchen, eine beliebige positiv definite quadratische Form durch die Wahl einer geeigneten Basis auf eine möglichst einfache Form zu bringen. Die denkbar einfachste Form ist

$$q(x, y) = ax^2 + cy^2,$$

wobei man die Bezeichnungen so wählen kann, dass $0 < a \leq c$ ist. Dann gilt offensichtlich Folgendes.

(i) Für alle x und y, die nicht beide gleich Null sind, ist

$$q(1, 0) \leq q(x, y).$$

(ii) Für alle x und y mit der Eigenschaft $y \neq 0$ gilt

$$q(0, 1) \leq q(x, y).$$

Viele Formen lassen sich aber durch umkehrbare ganzzahlige Substitutionen nicht diagonalisieren.

Definition 39. *Eine positiv definite quadratische Form q in zwei Variablen heißt reduziert, wenn sie die obigen Eigenschaften (i) und (ii) hat.*

Übersetzt in die Sprache der Formen auf Moduln bedeutet das Folgendes.

Definition 40. *Gegeben sei eine quadratische Form q_0 auf einem freien Modul V vom Rang 2 über dem Ring $\mathbf{Z}$. Eine Basis $\mathbf{v}$, $\mathbf{w}$ von V heißt reduziert, wenn $\mathbf{v}$ am kürzesten unter allen von $\mathbf{0}$ verschiedenen Vektoren ist und wenn $\mathbf{w}$ am kürzesten unter allen Vektoren ist, die kein Vielfaches von $\mathbf{v}$ sind.*

Für praktische Rechnungen ist die Sprache der Polynome zweckmäßiger.

Lemma 8. *Eine positiv definite quadratische Form in der Darstellung* (10) *ist genau dann reduziert, wenn*

$$|b| \le a \le c.$$

Beweis. Angenommen, q ist reduziert. Dann gilt

$$q(1,0) \le q(0,1), \qquad q(0,1) \le q(1,\pm 1),$$

also

$$a \le c, \qquad c \le a \pm b + c.$$

Subtrahieren wir bei der zweiten Gleichung $\pm b + c$ auf beiden Seiten, so erhalten wir die im Lemma angegebenen Ungleichungen.

Umgekehrt sei q eine positiv definite Form, deren Koeffizienten diesen Ungleichungen genügen, und x, y ganze Zahlen. Für $y \neq 0$ ist

$$x^2 \pm xy + y^2 = \frac{1}{2}\big(x^2 + y^2 + (x \pm y)^2\big) > 0,$$

also die ganzzahlige linke Seite sogar größer oder gleich 1, und es folgt

$$q(x,y) - c = ax^2 + bxy + c(y^2 - 1) \ge ax^2 - a|xy| + a(y^2 - 1) \ge 0.$$

Für $x \neq 0$ gilt offensichtlich

$$q(x,0) = ax^2 \ge a.$$

Daraus folgt, dass q reduziert ist. □

Satz 21. *Jede positiv definite quadratische Form in zwei Variablen mit ganzzahligen Koeffizienten ist äquivalent zu einer reduzierten Form.*

Beweis. Man kann eine gegebene positiv definite quadratische Form schrittweise durch lineare Substitutionen reduzieren, oder mit anderen Worten eine gegebene Basis in eine reduzierte Basis überführen. Wir beginnen mit der Substitution

$$\begin{aligned} x &= \quad y' \\ y &= x' + uy' \end{aligned}$$

welche q in die Form q' mit den Koeffizienten

$$a' = c, \qquad b' = b + 2cu, \qquad c' = a + bu + cu^2$$

überführt. Können wir erreichen, dass $b' = 0$ ist, so erhalten wir eine Diagonalform q'. Dazu müsste die ganze Zahl u gleich der rationalen Zahl $-\frac{b}{2c}$ sein, was im Allgemeinen unmöglich ist. Wir können aber die ganze Zahl u so wählen, dass

$$\left| u + \frac{b}{2c} \right| \leq \frac{1}{2}.$$

Dann ist $|b'| \leq c = a'$, d. h. eine der Ungleichungen aus Lemma 8 ist bereits erfüllt. Ist auch die andere Ungleichung $a' \leq c'$ erfüllt, so ist q' reduziert. Andernfalls ist $c' < a' = c$, und wir führen einen weiteren Reduktionsschritt aus. Auf diese Weise erhalten wir positiv definite quadratische Formen q', q'', q''', $q^{(4)}$ usw. Es kann nicht für alle i gelten $c^{(i)} < c^{(i-1)}$, weil die Koeffizienten $c^{(i)}$ für alle i nicht negativ sind. Also muss es ein i geben, so dass $c^{(i)} \geq c^{(i-1)} = a^{(i)}$, und dann ist $q^{(i)}$ reduziert. □

Wenden wir die Reduktion auf eine bereits reduzierte Form an, so können wir $u = 0$ wählen, und nach zwei Schritten erhalten wir die ursprüngliche Form.

In der Sprache der Basen bedeutet die Reduktion das Folgende. Man beginnt mit einem primitiven[11] Vektor $\mathbf{v}_1$ (im vorliegenden Fall $\mathbf{w}$) und wählt unter allen Vektoren, die kein Vielfaches von $\mathbf{v}_1$ sind, einen kürzesten Vektor $\mathbf{v}_2$. Dann wählt man unter allen Vektoren, die kein Vielfaches von $\mathbf{v}_2$ sind, einen kürzesten Vektor $\mathbf{v}_3$ usw. Für alle i bilden $\mathbf{v}_i$, $\mathbf{v}_{i+1}$ eine Basis, und für genügend große i sind diese Basen, eventuell nach Vertauschung der Vektoren, reduziert.

Beispiel. Wir betrachten die quadratische Form

$$q(x, y) = 36x^2 - 43xy + 13y^2.$$

Die nächste ganze Zahl bei $-\frac{b}{2c} = \frac{43}{26}$ ist $u = 2$, und wir erhalten

$$a' = 13, \qquad b' = -43 + 2 \cdot 13 \cdot 2 = 9, \qquad c' = 36 - 43 \cdot 2 + 13 \cdot 2^2 = 2.$$

Die Form

$$q'(x', y') = 13x'^2 + 9x'y' + 2y'^2$$

ist noch nicht reduziert, also fahren wir fort. Die nächste ganze Zahl bei $-\frac{b'}{2c'} = -\frac{9}{4}$ ist $u' = -2$, und wir erhalten

$$a'' = 2, \qquad b'' = 9 - 2 \cdot 2 \cdot 2 = 1, \qquad c'' = 13 - 9 \cdot 2 + 2 \cdot 2^2 = 3.$$

[11] Ein Vektor heißt reduziert, wenn er kein positives Vielfaches eines anderen Vektors ist.

Nun haben wir die reduzierte Form

$$q''(x'', y'') = 2x''^2 + x''y'' + 3y''^2$$

erhalten.

Satz 22. *Sind zwei verschiedene reduzierte positiv definite quadratische Formen äquivalent, so handelt es sich um*

$$ax^2 + bxy + cy^2 \qquad \textit{und} \qquad ax^2 - bxy + cy^2.$$

Beweis. Angenommen, q und q' sind positiv definit, reduziert und zueinander äquivalent. Wir können annehmen, dass $a \geq a'$. In den obigen Bezeichnungen ist $a' = q(r, s)$ für ganze Zahlen r und s. Wegen der Reduziertheit von q folgt

$$a \geq ar^2 + brs + cs^2 \geq a(r^2 + s^2) - a|rs| \geq a|rs|.$$

Wegen $ru - st = \pm 1$ können r und s nicht beide gleich Null sein.

Erster Fall: $|rs| = 1$. Dann ist $a = a' = a + brs + c$, also $|b| = c$, und wegen der Reduziertheit $a = c$. Nun ist

$$b' = p(r, s, t, u) = a(2rt - rs(ru + st) + 2su) = a(rt + su),$$

und die Zahl $|rt+su| = |(rt+su)rs| = |st+ru|$ kann wegen der Reduziertheit von q' höchstens 1 sein. Sie muss wegen $st+ru = 2st+1$ ungerade sein. Somit ist $|b'| = a$.

Zweiter Fall: $s = 0$. Dann ist $|r| = |u| = 1$, $a' = a$,

$$b' = p(r, 0, t, u) = (2at + bu)r.$$

Wegen der Reduziertheit von q und q' muss $|b| \leq a$ und $|b'| \leq a'$ sein. Mit der Dreiecksungleichung folgt

$$|2at| \leq |2at + bu| + |bu| \leq 2a,$$

also $|t| \leq 1$. Ist $t = 0$, so gilt $|b'| = |b|$, und ist $t = \pm 1$, so gilt in den letzten Ungleichungen das Gleichheitszeichen, also $|b| = a$ und $|b'| = a'$.

Dritter Fall: $r = 0$. Dann ist $|s| = |t| = 1$, $a' = c$, und wegen der Reduziertheit von q folgt $a = c$. Außerdem gilt

$$b' = p(0, s, t, u) = (bt + 2cu)s.$$

Wieder muss $|b| \leq a$ und $|b'| \leq a'$ sein, so dass

$$|2cu| \leq |bt + 2cu| + |bt| \leq 2c,$$

also $|u| \leq 1$. Ist $u = 0$, so folgt $|b'| = |b|$, und ist $u = \pm 1$, so gilt in den letzten Ungleichungen das Gleichheitszeichen, also $|b| = a$ und $|b'| = a'$.

Mit $d = d'$ folgt in allen drei Fällen $c = c'$. □

Nun kann man für zwei gegebene positiv definite quadratische Formen in zwei Variablen feststellen, ob sie äquivalent sind. Dazu ist zunächst einmal notwendig, dass sie die selbe Determinante haben. Wenn ja, sollte man beide reduzieren und mit Hilfe von Satz 22 prüfen, ob die reduzierten Formen äquivalent sind.

Nicht jede ganze Zahl kann als Diskriminante einer quadratischen Form in zwei Variablen auftreten. Wenn es eine quadratische Form mit Diskriminante d gibt, so ist d von der Form $4k$ oder $4k+1$ mit $k \in \mathbf{Z}$. Die Zahl b hat nämlich eine der Formen $4l$, $4l+1$, $4l+2$ oder $4l+3$, und d ist dann $4(4l^2 - ac)$, $4(4l^2 + 2l - ac) + 1$, $4(4l^2 + 4l + 1 - ac)$ bzw. $4(4l^2 + 6l + 2 - ac) + 1$.

Folgerung 15. *Es gibt nur endlich viele Äquivalenzklassen von positiv definiten quadratischen Formen mit vorgegebener Diskriminante.*

Beweis. Nach Satz 21 enthält jede Äquivalenzklasse eine reduzierte Form, also ist die Anzahl der Äquivalenzklassen mit Diskriminante d nicht größer als die Anzahl der reduzierten Formen mit Diskriminante d. Für eine reduzierte Form mit Diskriminante d gilt

$$4b^2 \leq 4ac = b^2 - d,$$

also $3b^2 \leq -d$. Somit gibt es nur endlich viele Möglichkeiten für b, und die Zahl $b^2 - d$ kann nur auf endlich viele Arten als $4ac$ dargestellt werden. □

Ist n eine ganze Zahl, so erhalten wir aus einer quadratischen Form q eine Form q', indem wir $q'(x, y) = nq(x, y)$ setzen. Da alle Eigenschaften von q' leicht aus denen von q folgen, kann man sich eigentlich auf das Studium von ganzzahligen quadratischen Formen beschränken, die nicht Vielfaches einer anderen Form sind. Solche nennt man primitive Formen; sie sind dadurch charakterisiert, dass ihre Koeffizienten teilerfremd sind. Insbesondere wird der Fall negativ definiter Formen auf den Fall positiv definiter Formen zurückgeführt.

Folgerung 15 gilt auch für indefinite Formen und wird für jene mit einer anderen Art von Reduktion bewiesen, die wir hier aber nicht behandeln werden. Beide Arten von Reduktion wurden von Carl Friedrich Gauß (1777–1855) behandelt. Es gibt auch eine Reduktionstheorie für positiv definite Formen in einer beliebigen Zahl von Variablen, die von Charles Hermite (1822–1901) und Hermann Minkowski (1864–1909) entwickelt wurde.

19 Spezielle Relativitätstheorie

Um diese Theorie zu motivieren, beginnen wir mit der Vorgeschichte. Die Kinematik ist das einfachste Teilgebiet der Physik, das die Bewegung von

Körpern beschreibt, ohne auf die Ursachen der Bewegung einzugehen. Traditionell stellt man sich das Weltall als dreidimensionalen Euklidischen Raum E vor. Auf dem Raum V der Verschiebungen von E ist ein Skalarprodukt s mit der zugehörigen Norm gegeben, und der Abstand zwischen zwei Punkten P und Q ist[12]

$$d(P,Q) = \left\|\overrightarrow{PQ}\right\|.$$

Die Zeitachse stellt man sich als eindimensionalen reellen affinen Raum T vor. Ist U der Raum der Verschiebungen der Zeitachse T, so ist für jeden von $\mathbf{0}$ verschiedenen Vektor festgelegt, ob er in die Vergangenheit oder die Zukunft zeigt. Eine Basis von U besteht aus einem Vektor $\mathbf{u}$, den wir in die Zukunft zeigen lassen. Dann ist durch die Festlegung $l(t\mathbf{u}) = t$ für $t \in \mathbf{R}$ eine Linearform l auf U definiert. Die Dauer von einem Zeitpunkt $t_1 \in T$ bis zu einem Zeitpunkt $t_2 \in T$ ist dann[13] $l\big(\overrightarrow{t_1 t_2}\big)$ und kann auch negativ sein.

Ein Ereignis geschieht zu einem Zeitpunkt t an einem Ort P, ist also durch ein geordnetes Paar (t, P) im vierdimensionalen affinen Raum $A = T \times E$ (der sogenannten Raumzeit) gegeben. Ordnen wir jedem Ereignis seinen Ort zu, so erhalten wir eine affine Abbildung $o : A \to E$. Ordnen wir jedem Ereignis seinen Zeitpunkt zu, so erhalten wir eine Abbildung $z : A \to T$.

Nun wollen wir die Bewegung eines Teilchens beschreiben. Bezeichnen wir seinen Ort zum Zeitpunkt $t \in T$ mit $f(t)$, so erhalten wir eine Abbildung $f : T \to E$. Den Graphen dieser Abbildung, also die Menge

$$\{(t, f(t)) \mid t \in T\} \subset T \times E$$

bezeichnet man als Weltlinie. Wir wollen hier nur geradlinige gleichförmige Bewegungen betrachten, deren Weltlinien Geraden sind, auf denen die Abbildung z nicht konstant ist. Eine solche hat einen Geschwindigkeitsvektor $\mathbf{v} \in V$, so dass für alle Zeitpunkte t_1 und t_2 gilt

$$\overrightarrow{f(t_1)f(t_2)} = l\big(\overrightarrow{t_1 t_2}\big)\mathbf{v}$$

(Weg gleich Zeit mal Geschwindigkeit).

Befindet man sich in einem leeren Weltall, so kann man nach den Gesetzen der Newtonschen Mechanik nicht entscheiden, ob man sich geradlinig gleichförmig bewegt oder sich in Ruhe befindet. Das zeitunabhängige Weltall E und die oben beschriebene Abbildung $o : A \to E$ haben also keinen physikalischen Sinn. Man muss sich von der Beschreibung der Raumzeit

[12] Der Abstand ist genaugenommen $d(P,Q)$ multipliziert mit der durch s festgelegten Längeneinheit.

[13] Die Dauer ist genaugenommen diese Zahl multipliziert mit der durch $\mathbf{u}$ festgelegten Zeiteinheit.

als Kreuzprodukt $T \times E$ verabschieden. Statt dessen findet die Newtonsche Mechanik in einem vierdimensionalen affinen Raum A statt, wobei eine affine Abbildung $z : A \to T$ gegeben ist, die jedem Ereignis F seinen Zeitpunkt zuordnet. Wir sagen, dass zwei Ereignisse F und G gleichzeitig geschehen, wenn $z(F) = z(G)$ ist. Für jeden Zeitpunkt t bilden die Ereignisse zum Zeitpunkt t einen dreidimensionalen affinen Unterraum $E_t = \{F \in A \mid z(F) = t\}$ von A.

Dass z eine affine Abbildung ist, bedeutet, dass es eine lineare Abbildung z' vom Vektorraum W der Verschiebungen der Raumzeit A in den Vektorraum U der Zeitverschiebungen gibt, so dass für alle Ereignisse F und G gilt

$$z'\big(\overrightarrow{FG}\big) = \overrightarrow{z(F)z(G)}.$$

Die Vektoren $\mathbf{x} \in W$ mit der Eigenschaft $z'(\mathbf{x}) = 0$, die wir räumliche Verschiebungen nennen, bilden einen Unterraum V von W, und auf diesem ist ein Skalarprodukt s mit der zugehörigen Norm gegeben. Der räumliche Abstand ist nur zwischen gleichzeitigen Ereignissen F und G erklärt als

$$d(F, G) = \big\|\overrightarrow{FG}\big\|.$$

Bewegt sich ein Teilchen geradlinig und gleichförmig, so ist seine Weltlinie eine Gerade in der Raumzeit W, auf der die Abbildung z nicht konstant ist. Sie hat dann einen eindeutig bestimmten Richtungsvektor $\mathbf{w} \in W$ mit der Eigenschaft $z'(\mathbf{w}) = \mathbf{u}$, den wir Vierergeschwindigkeit nennen. Bewegen sich zwei Teilchen geradlinig gleichförmig mit den Vierergeschwindigkeiten $\mathbf{w}_1$ und $\mathbf{w}_2$, so nennt man $\mathbf{w}_2 - \mathbf{w}_1 \in V$ die Relativgeschwindigkeit des zweiten Teilchens bezüglich des ersten. In der klassischen Mechanik spielen nur Relativgeschwindigkeiten eine Rolle. Offensichtlich gelten folgende Aussagen:

(i) Hat das zweite Teilchen bezüglich des ersten die Relativgeschwindigkeit $\mathbf{v}$, so hat das erste Teilchen bezüglich des zweiten die Relativgeschwindigkeit $-\mathbf{v}$.

(ii) (Additionstheorem) Hat das zweite Teichen die Relativgeschwindigkeit $\mathbf{v}_1$ bezüglich des ersten und hat ein drittes Teilchen die Relativgeschwindigkeit $\mathbf{v}_2$ bezüglich des zweiten, so ist die Relativgeschwindigkeit des dritten Teilchens bezüglich des ersten gleich

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2.$$

Ende des neunzehnten Jahrhunderts mehrten sich die Hinweise, dass die relative Lichtgeschwindigkeit im Vakuum immer den selben Betrag hat, nämlich $c = 299\,792\,458\,\mathrm{m/s}$, egal ob man sich auf die Lichtquelle zu oder von

ihr weg bewegt. Dies widerspricht eklatant dem genannten Additionstheorem für Geschwindigkeiten und zeigt, dass die oben dargestellte relative Kinematik höchtens für kleine Geschwindigkeiten die Wirklichkeit annähernd richtig beschreibt.

Albert Einstein (1879–1955) fand mit seiner speziellen Relativitätstheorie den Ausweg. Man muss sich nicht nur vom absoluten Ort verabschieden, sondern auch von der absoluten Zeit, d. h. die Zeitachse T und die Abbildung $z : A \to T$ haben keinen physikalischen Sinn. Statt dessen wird der Raum W der Verschiebungen der Raumzeit A mit einer symmetrischen Bilinearform b der Signatur $(1, 3, 0)$ versehen. Die isotropen Vektoren sind genau die Vierergeschwindigkeiten von Lichtstrahlen. Den Raum A mit dieser Zusatzstruktur nennt man Minkowski-Raum. Vektoren, auf denen die Spezialisierung q von b nichtnegative Werte annimmt, heißen zeitartig, diejenigen mit negativen Werten von q raumartig. Wir sagen, dass zeitartige Vektoren $\mathbf{w}_1 \neq \mathbf{w}_2$ die selbe Zeitorientierung haben, wenn $b(\mathbf{w}_1, \mathbf{w}_2) > 0$ ist. Dies ist eine Äquivalenzrelation mit zwei Äquivalenzklassen. Vektoren in der einen Klasse bezeichnen wir als in die Zukunft, Vektoren in der anderen als in die Vergangenheit gerichtet. Als Vierergeschwindigkeiten von Weltlinien für geradlinige gleichförmige Bewegungen kommen nur zeitartige Vektoren in Frage, die in die Zukunft gerichtet sind.

Um festzustellen, ob zwei Ereignisse F und G gleichzeitig stattfinden, betrachten wir den Mittelpunkt M ihrer Verbindungsstrecke in A. Dies bedeutet, dass $\overrightarrow{FM} = \overrightarrow{MG}$. Eine naive Definition der Gleichzeitigkeit würde besagen, dass die von F und G gegenseitig ausgesandten Lichtstrahlen sich auf der Weltlinie von M treffen. Diese Weltlinie hängt aber von der Wahl einer Vierergeschwindigkeit $\mathbf{w}$ ab, und das Selbe gilt somit für den Begriff der Gleichzeitigkeit. Bezeichnen wir das Ereignis des Zusammentreffens der Lichtstrahlen mit N, so bedeutet die Isotropie der Lichtstrahlen, dass

$$q(\overrightarrow{FM} + \overrightarrow{MN}) = 0, \qquad q(\overrightarrow{GM} + \overrightarrow{MN}) = 0.$$

Subtrahieren wir beide Gleichungen voneinander, so erhalten wir unter Beachtung von $\overrightarrow{GM} = -\overrightarrow{FM}$, dass

$$b(\overrightarrow{FM}, \overrightarrow{MN}) = 0.$$

Wir sehen also, dass Ereignisse F und G genau dann gleichzeitig bezüglich der Vierergeschwindigkeit $\mathbf{w}$ stattfinden, wenn

$$b(\overrightarrow{FG}, \mathbf{w}) = 0.$$

Bezeichnen wir das Orthogonalkomplement des von $\mathbf{w}$ erzeugten Unterraums mit $V_{\mathbf{w}}$, so können wir die Gleichzeitigkeit von zwei Ereignissen F und G bezüglich $\mathbf{w}$ durch die Bedingung $\overrightarrow{FG} \in V_{\mathbf{w}}$ ausdrücken.

Ist $q(\mathbf{w}) > 0$, so ist die Einschränkung von $-q$ auf $V_{\mathbf{w}}$ ein Skalarprodukt. Wir bezeichnen es mit $s_{\mathbf{w}}$ und kennzeichnen auch die zugehörige Norm mit einem Index $\mathbf{w}$. Für alle zu M gleichzeitigen Ereignisse F, von denen die Lichtstrahlen im Augenblick N eintreffen, gilt $q(\overrightarrow{FM}) + q(\overrightarrow{MN}) = 0$, also

$$\|\overrightarrow{FM}\|_{\mathbf{w}} = \sqrt{q(\overrightarrow{MN})}.$$

Ist die Dauer von M bis N gleich t, so haben die Strahlen nach unseren physikalischen Annahmen die Entfernung ct zurückgelegt. Es liegt also nahe, die Dauer von einem Ereignis M bis zu einem Ereignis N längs einer geraden Weltlinie durch

$$\frac{1}{c}\sqrt{q\big(\overrightarrow{MN}\big)}$$

zu definieren. Dann folgt, dass der Abstand zwischen zwei Ereignissen F und G, die bezüglich der Vierergeschwindigkeit $\mathbf{w}$ gleichzeitig stattfinden, gleich

$$\|\overrightarrow{FM}\|_{\mathbf{w}} = \sqrt{-q\big(\overrightarrow{FM}\big)}$$

ist. Man beachte, dass rechts kein $\mathbf{w}$ mehr auftaucht. In der Tat findet man für beliebige Ereignisse F und M mit der Eigenschaft $q\big(\overrightarrow{FM}\big) < 0$ eine Vierergeschwindigkeit, bezüglich derer sie gleichzeitig stattfinden, während Ereignisse M und N mit der Eigenschaft $q\big(\overrightarrow{MN}\big) \geq 0$ auf einer Weltlinie liegen. Insbesondere sieht man, dass auf einem Lichtstrahl die Zeit stehen bleibt.

Wir kommen nun zum Begriff der Relativgeschwindigkeit. Gegeben seien zwei Weltlinien mit Vierergeschwindigkeiten $\mathbf{w}_1$ und $\mathbf{w}_2$, wobei $q(\mathbf{w}_1) > 0$. Wir normieren $\mathbf{w}_1$ so, dass $q(\mathbf{w}_1) = c^2$, und schreiben $\mathbf{w}_2 = a\mathbf{w}_1 + \mathbf{v}$, wobei $\mathbf{v} \in V_{\mathbf{w}_1}$. Nun ist $ac^2 = b(\mathbf{w}_1, \mathbf{w}_2) > 0$, also können wir $\mathbf{w}_2$ so normieren, dass $a = 1$. Mit dieser Normierung bezeichnen wir $\mathbf{v} = \mathbf{w}_2 - \mathbf{w}_1$ als Relativgeschwindigkeit von $\mathbf{w}_2$ bezüglich $\mathbf{w}_1$. Wir erhalten $0 \leq q(\mathbf{w}_1 + \mathbf{v}) = q(\mathbf{w}_1) + q(\mathbf{v})$, also $\|\mathbf{v}\|_{\mathbf{w}_1} \leq c$, wobei genau dann Gleichheit gilt, wenn $\mathbf{w}_2$ isotrop, also die zweite Weltlinie ein Lichtstrahl ist.

Satz 23. *(i) Ist $\mathbf{v}$ die Relativgeschwindigkeit von $\mathbf{w}_2$ bezüglich $\mathbf{w}_1$ und $\mathbf{u}$ die Relativgeschwindigkeit von $\mathbf{w}_1$ bezüglich $\mathbf{w}_2$, so gilt $\|\mathbf{v}\|_{\mathbf{w}_1} = \|\mathbf{u}\|_{\mathbf{w}_2}$.*

(ii) (Additionstheorem) Ist $\mathbf{v}_1$ die Relativgeschwindigkeit von $\mathbf{w}_1$ und $\mathbf{v}_2$ die Relativgeschwindigkeit von $\mathbf{w}_2$, beide bezüglich einer Vierergeschwindigkeit $\mathbf{w}$, so gilt für den Betrag v der Relativgeschwindigkeit zwischen $\mathbf{w}_1$ und $\mathbf{w}_2$

$$v \leq \frac{\|\mathbf{v}_1 - \mathbf{v}_2\|_{\mathbf{w}}}{1 - s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2)/c^2},$$

wobei genau dann Gleichheit gilt, wenn $\mathbf{v}_1$ *und* $\mathbf{v}_2$ *kollinear sind.*

Beweis. Es gilt

$$q(\mathbf{w}_1) = c^2, \qquad q(\mathbf{w}_2) = c^2 - \|\mathbf{v}\|_{\mathbf{w}_1}^2, \qquad b(\mathbf{w}_1, \mathbf{w}_2) = c^2.$$

Daraus folgt

$$1 - \frac{\|\mathbf{v}\|_{\mathbf{w}_1}}{c^2} = \frac{q(\mathbf{w}_1)q(\mathbf{w}_2)}{b(\mathbf{w}_1, \mathbf{w}_2)^2}.$$

Die rechte Seite bleibt unverändert, wenn wir $\mathbf{w}_1$ und $\mathbf{w}_2$ vertauschen, und Behauptung (i) folgt.

Die rechte Seite bleibt auch unverändert, wenn wir $\mathbf{w}_1$ oder $\mathbf{w}_2$ mit einer reellen Zahl multiplizieren. Die Formel gilt also auch ohne die beschriebene Normierung. Insbesondere können wir $\mathbf{w}_1$ und $\mathbf{w}_2$ in Bezug auf eine Vierergeschwindigkeit $\mathbf{w}$ normieren, so dass $\mathbf{w}_1 = \mathbf{w} + \mathbf{v}_1$, $\mathbf{w}_2 = \mathbf{w} + \mathbf{v}_2$. Setzen wir die Ausdrücke

$$q(\mathbf{w}_1) = c^2 - \|\mathbf{v}_1\|_{\mathbf{w}}^2, \quad q(\mathbf{w}_2) = c^2 - \|\mathbf{v}_2\|_{\mathbf{w}}^2, \quad b(\mathbf{w}_1, \mathbf{w}_2) = c^2 - s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2)$$

ein, so erhalten wir

$$1 - \frac{v^2}{c^2} = \frac{(c^2 - \|\mathbf{v}_1\|_{\mathbf{w}}^2)(c^2 - \|\mathbf{v}_2\|_{\mathbf{w}}^2)}{(c^2 - s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2))^2}.$$

Umstellen ergibt

$$\begin{aligned}
\frac{v^2}{c^2} &= \frac{(c^2 - s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2))^2 - (c^2 - \|\mathbf{v}_1\|_{\mathbf{w}}^2)(c^2 - \|\mathbf{v}_2\|_{\mathbf{w}}^2)}{(c^2 - s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2))^2} \\
&= \frac{s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2)^2 - \|\mathbf{v}_1\|_{\mathbf{w}}^2\|\mathbf{v}_2\|_{\mathbf{w}}^2 + c^2(\|\mathbf{v}_1\|_{\mathbf{w}}^2 - 2s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2) + \|\mathbf{v}_2\|_{\mathbf{w}}^2)}{(c^2 - s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2))^2} \\
&\leq \frac{c^2(\|\mathbf{v}_1\|_{\mathbf{w}}^2 - 2s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2) + \|\mathbf{v}_2\|_{\mathbf{w}}^2)}{(c^2 - s_{\mathbf{w}}(\mathbf{v}_1, \mathbf{v}_2))^2},
\end{aligned}$$

wobei wir im letzten Schritt die Cauchy-Schwarz-Ungleichung benutzt haben, in der genau dann Gleichheit gilt, wenn $\mathbf{v}_1$ und $\mathbf{v}_2$ kollinear sind. □

Satz 24 (Längenkontraktion)**.** *Sind zwei Weltlinien mit der selben Vierergeschwindigkeit* $\mathbf{w}$ *um den Vektor* $\mathbf{x} \in V_{\mathbf{w}}$ *versetzt und hat die Vierergeschwindigkeit* $\mathbf{w}'$ *die Relativgeschwindigkeit* $\mathbf{v}$ *bezüglich* $\mathbf{w}$*, so ist der Abstand der Teilchen bezüglich* $\mathbf{w}'$ *gleich*

$$\sqrt{\|\mathbf{x}\|_{\mathbf{w}}^2 - s_{\mathbf{w}}(\mathbf{x}, \mathbf{v})^2/c^2}.$$

Sind $\mathbf{x}$ *und* $\mathbf{v}$ *kollinear, so erhalten wir*

$$\|\mathbf{x}\|_{\mathbf{w}}\sqrt{1 - \|\mathbf{v}\|_{\mathbf{w}}^2/c^2}.$$

Beweis. Angenommen, F und G sind gleichzeitige Positionen der beiden Teilchen bezüglich $\mathbf{w}$. Dann ist $\mathbf{x} = \overrightarrow{FG}$. Liegt G' auf der Weltlinie des zweiten Teilchens, so ist $\mathbf{x}' = \overrightarrow{FG'} = \mathbf{x} + t\mathbf{w}$ für eine Zahl t. Wir wählen t so, dass G' bezüglich $\mathbf{w}'$ gleichzeitig mit F stattfindet, also

$$0 = b(\mathbf{x}', \mathbf{w}') = b(\mathbf{x}, \mathbf{w}') + tb(\mathbf{w}, \mathbf{w}').$$

Wir normieren $q(\mathbf{w}) = c^2$ und $\mathbf{w}' = \mathbf{w} + \mathbf{v}$ mit $\mathbf{v} \in V_{\mathbf{w}}$. Nun ist $b(\mathbf{w}, \mathbf{w}') = c^2$ und $tc^2 = -b(\mathbf{x}, \mathbf{w}')$. Wegen $b(\mathbf{x}, \mathbf{w}) = 0$ folgt $b(\mathbf{x}, \mathbf{w}') = b(\mathbf{x}, \mathbf{v})$ sowie

$$q(\mathbf{x}') = q(\mathbf{x}) + t^2 q(\mathbf{w}) = q(\mathbf{x}) + b(\mathbf{x}, \mathbf{v})^2 / c^2.$$

Mit der Definition von $s_{\mathbf{w}}$ folgt die Behauptung. □

Satz 25 (Zeitdilatation)**.** *Es seien M und N Eregnisse auf einer Weltlinie mit der Vierergeschwindigkeit $\mathbf{w}$ sowie M' und N' Ereignisse auf einer Weltlinie mit der Vierergeschwindigkeit $\mathbf{w}'$, die zu M bzw. N bezüglich $\mathbf{w}$ gleichzeitig stattfinden. Weiter sei t die Dauer von M bis N und t' die Dauer von M' bis N'. Ist v der Betrag der Relativgeschwindigkeit zwischen $\mathbf{w}$ und $\mathbf{w}'$, so ist*

$$t' = t\sqrt{1 - v^2/c^2}.$$

Beweis. Normieren wir $\mathbf{w}$ durch $q(\mathbf{w}) = c^2$, so gilt $\overrightarrow{MN} = t\mathbf{w}$, und normieren wir $\mathbf{w}'$ durch $\mathbf{w}' = \mathbf{w} + \mathbf{v}$ mit $\mathbf{v} \in V_{\mathbf{w}}$, so gilt $\overrightarrow{M'N'} = \frac{ct'}{\sqrt{q(\mathbf{w}')}}\mathbf{w}'$. Es folgt

$$\overrightarrow{NN'} - \overrightarrow{MM'} = \frac{ct'}{\sqrt{q(\mathbf{w}')}}\mathbf{w}' - t\mathbf{w}.$$

Da die linke Seite in $V_{\mathbf{w}}$ liegt, ist

$$\frac{ct' b(\mathbf{w}', \mathbf{w})}{\sqrt{q(\mathbf{w}')}} = tq(\mathbf{w}).$$

Wegen $b(\mathbf{v}, \mathbf{w}) = 0$ gilt $b(\mathbf{w}', \mathbf{w}) = q(\mathbf{w})$ und

$$q(\mathbf{w}') = q(\mathbf{w}) + q(\mathbf{v}).$$

Nun folgt die Behauptung mit $q(\mathbf{v}) = -v^2$. □